Optimal maintenance policies for a safety-critical system and its deteriorating sensor

Optimal Maintenance Policies for a Safety-Critical System and Its Deteriorating Sensor

Chiel van Oosterom[a], Lisa M. Maillart[b], Jeffrey P. Kharoufeh[b]

[a] Econometric Institute, Erasmus University Rotterdam
PO Box 1738, 3000 DR Rotterdam, The Netherlands and
School of Industrial Engineering, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[b] Department of Industrial Engineering, University of Pittsburgh
1025 Benedum Hall, 3700 O'Hara Street, Pittsburgh, PA 15261, USA

September 3, 2017

Abstract

We consider the integrated problem of optimally maintaining an imperfect, deteriorating sensor and the safety-critical system it monitors. The sensor's costless observations of the binary state of the system become less informative over time. A costly full inspection may be conducted to perfectly discern the state of the system, after which the system is replaced if it is in the out-of-control state. In addition, a full inspection provides the opportunity to replace the sensor. We formulate the problem of adaptively scheduling full inspections and sensor replacements using a partially observable Markov decision process (POMDP) model. The objective is to minimize the total expected discounted costs associated with system operation, full inspection, system replacement, and sensor replacement. We show that the optimal policy has a threshold structure and demonstrate the value of coordinating system and sensor maintenance via numerical examples.

Keywords: Maintenance optimization; Sensor deterioration; Partially observable Markov decision process; Threshold policy

1 Introduction

Failures of safety-critical systems, such as those encountered in chemical plants, hospitals, or nuclear power reactors, can pose risks to human life, harm the environment, or lead to substantial economic losses. The operation of these systems is typically subject to strict requirements based

on quantitative risk assessments and regulatory compliance standards (Fowler [9]). If a system is operated in an out-of-control state such that these requirements are not met, then it is overexposed to undue risks. To achieve maximum safety, it is essential that the state of the system be known at any moment in time. Then, as soon as the system transitions into the out-of-control state, maintenance can be performed in order to restore it to the in-control state. However, carrying out a site visit to conduct a full inspection of the system, which may also necessitate an interruption of operations, is usually expensive and cannot be done on a frequent basis. To help alleviate this problem, sensors (e.g., accelerometers, thermocouples, or wear debris sensors) are often deployed to obtain imperfect measurements of the state of the system (Davies [8]).

The informativeness of a sensor's measurements determines its ability to support maintenance decisions; naturally, better maintenance decisions can be made if the sensor provides more reliable measurements of the state of the system. Therefore, along with cost, measurement quality is a primary consideration when choosing the most appropriate condition monitoring technology. Srinivasan and Parlikad [38] develop a method to assess and compare the value of different condition monitoring techniques for infrastructure assets, and illustrate their approach using a case example. They quantify the value of condition monitoring as the benefit of having imperfect information over having no information. In applications where partial information about the state of the system can also be inferred in the absence of any sensor technology, such as a production system whose output is subjected to quality control procedures, a different evaluation of the value of condition monitoring may be used (cf. Gilbert and Bar [11]). To enhance the measurement quality of a condition monitoring system, it is also possible to combine multiple types of sensors or to add multiple redundant sensors that all measure the same system parameters (see Ray and Phoha [33]).

The implicit assumption in existing maintenance optimization models is that, once a condition monitoring system has been selected and the sensors have been implemented, the performance of sensors remains constant over time, i.e., the quality of the measurements of the state of the system is considered to be stationary. However, as evidenced by durability studies of piezo wafer active sensor systems (Blackshire et al. [3]) and fiber-optic strain sensors (Habel and Bismarck [14]), for example, sensors may also deteriorate. Therefore, to fully realize the benefits of condition-based maintenance strategies, it is necessary to consider sensor maintenance (e.g., via replacement or recalibration). Coble et al. [6] report that for nuclear power plants in the United States, periodic recalibration of all safety-related sensors is mandatory. Such settings motivate us to explore joint optimal system and sensor maintenance policies.

1.1 Problem Description and Contributions

In this paper, we investigate the simultaneous maintenance of a safety-critical system and the deteriorating sensor that monitors the system. We shall use the term "system" throughout this paper,

although in reality the sensor might be monitoring a component of a safety-critical system (e.g., a safety valve in a nuclear power plant). The system's binary state (in-control or out-of-control) evolves as a discrete-time Markov chain. The sensor provides costless, imperfect observations of the state of the system; however, due to deterioration, the sensor provides less informative observations as it ages. Following each sensor observation, a costly full inspection may be conducted to perfectly discern the state of the system. If the system is found to be in the out-of-control state, then it is restored to the in-control state by a system replacement. Moreover, a full inspection offers an opportunity to replace the sensor, so as to acquire more informative future observations of the state of the system. We seek to determine how to adaptively schedule full inspections and sensor replacements to minimize the total expected discounted costs due to system operation, full inspection, system replacement, and sensor replacement. To this end, we formulate the problem using a partially observable Markov decision process (POMDP) model.

The main result of this paper is a characterization of the structure of the optimal policy, which is shown to be a threshold policy. Specifically, there exists (*i*) a sensor-age-dependent threshold such that it is optimal to perform a full inspection if and only if the probability that the system is in the out-of-control state exceeds that threshold, and — if it is ever optimal to replace the sensor — (*ii*) a threshold such that, at the time of a full inspection, sensor replacement is optimal if and only if the sensor age exceeds the threshold. The sensor-age-dependent threshold on the probability that the system is in the out-of-control state is nonincreasing in the range of sensor ages for which sensor replacement is optimal at a full inspection. When applied to the special case of a non-deteriorating sensor, our structural results generalize existing results in that we impose no assumptions on the transition probability matrix and only general assumptions on the cost parameters.

By way of numerical examples, we compare the optimal policy with heuristic policies to gain more insight into the value of coordinating system and sensor maintenance. These examples illustrate that using a constant threshold on the probability that the system is in the out-of-control state, as opposed to the optimal sensor-age-dependent threshold, can simplify the policy structure at a relatively small increase in cost (1.0% and 1.5%). When the coordination between system and sensor maintenance is relaxed further by simply periodically replacing the sensor irrespective of the information available on the state of the system, the optimality gap can be much larger (11.9% and 14.8%).

1.2 Related Literature

By formulating our problem of optimally maintaining a safety-critical system and its deteriorating sensor using a POMDP model, we follow a common approach in the literature on maintenance optimization problems with imperfect information on the state of the system. Researchers have developed and analyzed POMDP models for a number of ways in which information may be im-

perfect. There exist models in which costless, imperfect observations of the state of the system are made at every decision epoch (White [41], Ghasemi et al. [10], Grosfeld-Nir [13]) and models in which an observation of the state of the system is only made at a decision epoch if it is decided to conduct a costly inspection (Ross [36], Rosenfield [35], Maillart and Zheltova [26]). The combination of costless, imperfect observations and costly observations is considered in White [40] and Ohnishi et al. [30]. In all papers listed above, it is assumed that the probabilistic relation between the state of the system and the observation that is obtained, which in POMDP models is described by an (action-dependent) observation matrix, is stationary over time. There are also some recent works that include the possibility of choosing an inspection type at an inspection, where each inspection type is associated with a different cost and informativeness (e.g., a menu of different tests is available, or a choice can be made between full inspection and sampling; see Kuo [20], Maillart et al. [27], Kim and Makis [18]). The assumption in these papers is that for each inspection type the probabilistic relation between the state of the system and the observation that is obtained is stationary over time, and moreover, the same inspection types are available at every decision epoch. The distinguishing feature of our problem is that, because observations are obtained from a deteriorating sensor, the quality of observations is not stationary. To incorporate sensor deterioration, we propose a POMDP model in which the observation matrix depends on the age of the sensor. In this way, the informativeness of sensor observations can be modeled as dependent on the sensor age.

More specifically, to model the age-dependent informativeness of sensor observations, we draw upon the classic work of Blackwell [4, 5] on the comparison of experiments (for a detailed review on the comparison of experiments, the reader is referred to Le Cam [21]). Blackwell [4, 5] considers a single-stage decision problem in which a decision maker is to minimize her expected loss, where loss is a function of the action she chooses and the unknown state of nature. Before selecting an action, the decision maker, who has a prior belief about the state of nature expressed as a probability distribution on a set of candidate states, performs an experiment to gather additional information. This experiment is characterized by a collection of (known) probability distributions, one for each candidate state, defined over a set of possible outcomes. The actual probability distribution of the outcome of the experiment is the one associated with the true state of nature. After observing the outcome of the experiment, the decision maker updates her belief, and then selects an action to minimize the expected loss given her posterior belief. The minimum a priori expected loss for the decision problem is obtained by taking the expectation with respect to the outcome of the experiment. In this context, the question arises of how to compare two experiments defined for the same set of candidate states in terms of their informativeness. Several methods of comparison have been suggested. For example, one may define an experiment to be more informative than another experiment (a) if the former experiment yields a lower minimum a priori expected loss for all prior beliefs and all loss functions, or (b) if the outcomes of the latter experiment can be represented as

noise-corrupted outcomes of the former experiment (in which case the latter experiment is also called a "garbling" of the former). Blackwell [5] proves that the two methods above are equivalent, hence defining one method for ordering experiments according to their informativeness: the Blackwell order.

There are some major differences between POMDPs and the decision problem considered by Blackwell [4, 5]. A POMDP is a multi-stage decision problem in which the partially observable state changes dynamically over time, and actions do not only influence costs, but also the state transitions and the observations that are made. Yet, receiving an observation in a POMDP is similar to performing an experiment. In our POMDP model, the observation matrix associated with the age of the sensor specifies for both the in-control and the out-of-control state a probability distribution over possible values of the sensor observation; and upon receiving a sensor observation, the distribution of which corresponds to the true but unknown state of the system, the probability that the system is in the out-of-control state is updated. Recognizing this similarity, we adopt the Blackwell order to impose a relation between the observation matrices implied by different sensor ages such that sensor observations are less informative if the sensor is older.

The Blackwell order has been used by other researchers to capture informativeness relations in POMDP models. White and Harrington [42], Rieder [34], and Zhang [43] draw a comparison between POMDPs that have the same underlying Markov decision process (MDP) but a different observation process. They use the Blackwell order to show that the more informative observation process results in a lower total expected discounted cost. The Blackwell order is also employed to assist in the identification of dominated actions in POMDP models for such diverse applications as sensor scheduling problems (Krishnamurthy and Djonin [19]) and sequential hypothesis testing problems with various modes of gathering information (Naghshvar and Javidi [29]). Most related to our use of the Blackwell order are results in sequential hypothesis testing problems by Lévesque and Maillart [22], Ulu and Smith [39], and Alizamir et al. [2] that provide conditions under which the value of gathering additional information decreases over time. In our model, observations also become less informative over time; however, the option to reset the observation matrix at a cost (i.e., sensor replacement) introduces a trade-off between informativeness and cost that is present in none of the aforementioned works.

The remainder of the paper is organized as follows. In Section 2, we present our POMDP model for the problem of optimally maintaining a safety-critical system and its deteriorating sensor. Structural results on the optimal policy are established in Section 3, and numerical examples illustrating the value of coordinating system and sensor maintenance are given in Section 4. In Section 5, we conclude and suggest directions for further research.

2 Model Formulation

Consider a safety-critical system whose condition can be classified as either in-control '1' or out-of-control '2'. The system operates over an infinite horizon, over which time is divided into periods of equal length. The periods are indexed by the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The state of the system evolves as a discrete-time Markov chain over the state space $\mathcal{S} = \{1, 2\}$ according to transition probability matrix $P = [p_{ij}]$, $i, j \in \mathcal{S}$. In the out-of-control state, the system is exposed to an increased level of risk. This is quantified by a per-period cost $c_d > 0$ for operating the system in the out-of-control state. Note that c_d represents a penalty cost whose value is determined in a quantitative risk assessment.

At each transition epoch, a sensor provides a costless, imperfect observation of the state of the system, which takes a value in the finite observation space $\mathcal{O} = \{0, 1, \ldots, y\}$. To reflect the fact that the sensor deteriorates over time, the probabilistic relation between the state of the system and the observation is dependent on the sensor age. That is, the observation matrix $Q(t) = [q_{ik}(t)]$, $i \in \mathcal{S}$, $k \in \mathcal{O}$, is defined as a function of the sensor age: for sensor age $t \in \mathbb{N}_0$, $q_{ik}(t)$ gives the probability of observing $k \in \mathcal{O}$ if the state of the system is $i \in \mathcal{S}$. To further formalize that sensor deterioration decreases the information content of sensor observations, we assume the observation matrices are ordered according to the Blackwell order (Blackwell [4, 5]).

Definition 1 (Blackwell order)**.** Observation matrix Q is more informative than observation matrix $\widehat{Q}$, denoted by $Q \succeq_B \widehat{Q}$, if there exists a stochastic matrix X such that $\widehat{Q} = QX$.

We assume that $Q(t) \succeq_B Q(t+1)$ and denote by $X(t) = [x_{kl}(t)]$, $k, l \in \mathcal{O}$, the stochastic matrix such that $Q(t+1) = Q(t)X(t)$, for all $t \in \mathbb{N}_0$. Under this assumption, an observation by a sensor of age $t+1$ can be regarded as if it were an observation by a sensor of age t to which random noise has been added through a stochastic transformation by $X(t)$. Hence, the older the sensor, the more noise its observations contain.

Full inspections cost $c_s > 0$ and yield a perfect observation of the state of the system. If a full inspection identifies the system to be in the out-of-control state, it is restored to the in-control state by a replacement at additional cost $c_r \geq 0$. Both full inspection and system replacement take negligible time. Furthermore, a full inspection provides the opportunity for a sensor replacement; it may be decided to replace the sensor by a new one at cost $c_i > 0$. Sensor replacement is also assumed to take negligible time. All costs are discounted by a factor $\beta \in [0, 1)$ per period. Our aim is to schedule full inspections and sensor replacements to minimize the total expected discounted costs associated with system operation, full inspection, system replacement, and sensor replacement over an infinite horizon.

This problem can be modeled as a POMDP whose underlying state is the pair $(i, t) \in \mathcal{S} \times \mathbb{N}_0$, where i is the state of the system and t is the age of the sensor. Because the sensor observations

provide imperfect information about the system state, the underlying state is partially observable and cannot be directly used for decision making. Instead, a belief about the state of the system can serve, together with the (known) sensor age, as a basis for optimal decision making (cf. Monahan [28], Smallwood and Sondik [37]). The belief, which we express as the probability that the system is in the out-of-control state, is updated in a Bayesian manner as sensor observations are received. Following this approach, we cast the POMDP as an MDP on the information state space $\Omega = [0,1] \times \mathbb{N}_0$, where information state $(\pi, t) \in \Omega$ denotes that the system is in the out-of-control state with probability π and the sensor age is t. The action space is given by $\mathcal{A} = \{C, S, R\}$, where C denotes "continue operating," S denotes "full inspection without sensor replacement," and R denotes "full inspection with sensor replacement" with the understanding that full inspection also implies replacing the system if it is in the out-of-control state.

A policy is a rule that prescribes, for any information state, an action to be taken. Suppose at the start of a period, the information state is $(\pi, t) \in \Omega$. If action C is taken, the following sequence of events takes place:

1. An immediate expected cost πc_d is incurred;
2. The state of the system transitions according to P and the sensor age increments to $t+1$;
3. The sensor provides an observation of the state of the system. The probability of receiving observation $k \in \mathcal{O}$ is

$$\sigma(k; \pi, t) = ((1-\pi)p_{11} + \pi p_{21})q_{1k}(t+1) + ((1-\pi)p_{12} + \pi p_{22})q_{2k}(t+1);$$

4. Having received a sensor observation $k \in \mathcal{O}$, $\sigma(k; \pi, t) > 0$, Bayes' rule is used to update the probability that the system is in the out-of-control state to

$$\psi(\pi, t, k) = \frac{((1-\pi)p_{12} + \pi p_{22})q_{2k}(t+1)}{\sigma(k; \pi, t)}.$$

(We define $\psi(\pi, t, k) = 0$ for all $k \in \mathcal{O}$ such that $\sigma(k; \pi, t) = 0$.) This results in a new information state $(\psi(\pi, t, k), t+1)$ at the beginning of the next period.

Because full inspection, system replacement, and sensor replacement are instantaneous, the sequence of events after taking action S or R can be stated in terms of the sequence of events after taking action C. If action S is taken, an immediate expected cost $c_s + \pi c_r$ is incurred, and what follows is identical to when action C is taken in information state $(0, t)$. If action R is taken, an immediate expected cost $c_s + c_i + \pi c_r$ is incurred, and what follows is identical to when action C is taken in information state $(0, 0)$.

Let $\mathcal{V}$ denote the set of all bounded, real-valued functions on Ω, which we refer to as value functions. The optimal value function V^*, which gives the minimum total expected discounted cost as a function of the initial information state, is the unique solution to the optimality equations

$$V(\pi, t) = \min_{a \in \mathcal{A}} H^a(\pi, t, V) \tag{1}$$

for all $(\pi, t) \in \Omega$, where

$$H^C(\pi, t, V) = \pi c_d + \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) V(\psi(\pi, t, k), t+1),$$
$$H^S(\pi, t, V) = c_s + \pi c_r + \beta \sum_{k \in \mathcal{O}} \sigma(k; 0, t) V(\psi(0, t, k), t+1),$$
$$H^R(\pi, t, V) = c_s + c_i + \pi c_r + \beta \sum_{k \in \mathcal{O}} \sigma(k; 0, 0) V(\psi(0, 0, k), 1),$$

for all $(\pi, t) \in \Omega$ and $V \in \mathcal{V}$. The optimal policy, which we denote by δ^*, takes action $\delta^*(\pi, t) = \arg\min_{a \in \mathcal{A}} H^a(\pi, t, V^*)$ in information state (π, t), for all $(\pi, t) \in \Omega$.

Remark 1. This model excludes the possibility of conducting a site visit to replace the sensor without a system inspection. That is, we assume that a sensor replacement always comes with a perfect observation of the state of the system. This is certainly true if a sensor of age 0 yields perfect observations (i.e., for all $k \in \mathcal{O}$, there exists at most one system state $i \in \mathcal{S}$ such that $q_{ik}(0) > 0$), when one immediately discerns the state of the system through the sensor. However, even if a new sensor does not yield perfect observations, its observation might already contain enough information for a maintenance engineer who is present on-site to obtain a good indication of the state of the system.

Remark 2. We assume that the system is only replaced if a full inspection identifies the system to be in the out-of-control state. However, the cost structure also allows us to model scenarios in which the system is replaced regardless of the inspection outcome, or no inspection is performed before replacing the system. This special case can be examined by setting $c_r = 0$ and incorporating the cost of system replacement in c_s. The issue of which approach is more cost effective — performing full inspections at site visits or replacing the system without prior inspection — is not explored here. Likewise, we do not consider policies that decide dynamically, as a function of the information state, whether an inspection is performed at a site visit.

3 Structural Results

This section investigates the structural properties of the optimal policy for the POMDP model described in Section 2. In Section 3.1, we derive a necessary and sufficient condition to characterize the case in which it is optimal not to apply any maintenance actions (i.e., full inspections with or without sensor replacement) at all and give a closed-form expression for the associated optimal value function. For the alternative case where there exist information states in which it is optimal to perform a maintenance action, in Sections 3.2 and 3.3, we show monotonicity properties of the optimal value function and characterize the structure of the optimal policy. Finally, in Section 3.4, we consider the implications of our results when the sensor does not deteriorate. The proofs of the results are given in the Appendix.

3.1 Condition for the Optimality of No Maintenance

Intuitively, if the risks associated with operating the system in the out-of-control state are small relative to the costs of maintenance, it may not be economical to perform maintenance, and the sensor becomes irrelevant. In Theorem 1, we show that

$$(1-\beta(p_{22}-p_{12}))^{-1}c_d \leq c_s + c_r \tag{2}$$

is a necessary and sufficient condition for it to be optimal to continue operating in all information states. Also, we provide a closed-form expression for the resulting optimal value function.

Theorem 1. *Let the policy δ be defined by $\delta(\pi,t) = C$ for all $(\pi,t) \in \Omega$ and the value function V be defined by*

$$V(\pi,t) = \beta p_{12}(1-\beta)^{-1}(1-\beta(p_{22}-p_{12}))^{-1}c_d + \pi(1-\beta(p_{22}-p_{12}))^{-1}c_d \tag{3}$$

for all $(\pi,t) \in \Omega$. Then, $\delta = \delta^*$ and $V = V^*$ if and only if condition (2) holds.

From Theorem 1, it is seen that under condition (2), the optimal value function is constant in the sensor age, as decisions are not adapted to the sensor observations. Also, because $1-\beta(p_{22}-p_{12}) > 0$, it holds that the optimal value function is increasing in the probability that the system is in the out-of-control state.

3.2 Monotonicity of the Optimal Value Function

In this section, we derive results on the optimal value function when condition (2) does not hold, i.e.,

$$(1-\beta(p_{22}-p_{12}))^{-1}c_d > c_s + c_r. \tag{4}$$

Under condition (4), we cannot obtain a closed-form expression for the optimal value function, but we can establish results on its form. Our first result (Lemma 1) establishes monotonicity in the probability that the system is in the out-of-control state. We provide a non-negative lower bound on the change in the optimal value function as a function of an increase in the probability of the system being out of control.

Lemma 1. *If condition (4) holds, then $V^*(\hat{\pi},t) - V^*(\pi,t) \geq (\hat{\pi}-\pi)c_r$ for all $\pi,\hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$.*

A key observation in the proof of Lemma 1 (see Appendix) is that we may restrict our attention to problem instances where, for all $t \in \mathbb{N}_0$, the observation matrix $Q(t)$ is (i) totally positive of order two (TP_2), and (ii) such that, for all $k,l \in \mathcal{O}$ with $k \leq l$, $q_{2k}(t) > 0$ implies $q_{2l}(t) > 0$. The TP_2 property is defined as follows (Karlin [17]).

Definition 2. Observation matrix Q is TP_2 if $q_{il}q_{jk} \leq q_{ik}q_{jl}$ for all $i, j \in \mathcal{S}$ and $k, l \in \mathcal{O}$ such that $i \leq j$ and $k \leq l$.

These properties of the observation matrices ((i) and (ii)) guarantee a monotonic relationship between the current probability that the system is in the out-of-control state, the next sensor observation, and the updated probability that the system is in the out-of-control state (see Proposition 1 in the Appendix).

Remark 3. Any results on V^* or δ^* derived under conditions (i) and (ii) above apply to the general case where no assumptions are made on the observation matrices. This is true because, for every problem instance, there exists another problem instance that satisfies (i) and (ii) and is equivalent in the sense that V^* and δ^* are identical. The argument proceeds as follows. For a standard POMDP model, Lovejoy [23, p.739] points out that the optimal value function and the optimal policy are unaffected by a permutation of the columns of the observation matrix. Lovejoy's reasoning extends to our model, where the columns of $Q(t)$ may be permuted for all $t \in \mathbb{N}_0$. The crucial point is that the Blackwell order is preserved under column permutations: if, for $t \in \mathbb{N}_0$, $X(t)$ is a stochastic matrix such that $Q(t+1) = Q(t)X(t)$ and if $\widetilde{Q}(t)$ and $\widetilde{Q}(t+1)$ are column permutations of $Q(t)$ and $Q(t+1)$, then $\widetilde{X}(t)$ obtained by permuting the rows and columns of $X(t)$ accordingly is a stochastic matrix such that $\widetilde{Q}(t+1) = \widetilde{Q}(t)\widetilde{X}(t)$. The argument is complete by noting that because $|\mathcal{S}| = 2$, for every observation matrix $Q(t)$, $t \in \mathbb{N}_0$, there exists a column permutation that is TP_2 and such that, for all $k, l \in \mathcal{O}$ with $k \leq l$, $q_{2k}(t) > 0$ implies $q_{2l}(t) > 0$ — namely, a permutation such that, for some $l \in \mathcal{O}$, $q_{2k}(t) = 0$ for $k < l$ and $q_{1k}(t)/q_{2k}(t)$ is nonincreasing in k for $k \geq l$.

Our second result on the form of the optimal value function (Lemma 2) establishes monotonicity in the sensor age.

Lemma 2. *If condition (4) holds, then $V^*(\pi, t)$ is nondecreasing in t for all $\pi \in [0, 1]$.*

We note that the result of Lemma 2 holds under both conditions (2) and (4). However, under condition (2), a more specific result has already been established, namely a closed-form expression of the optimal value function in which $V^*(\pi, t)$ is constant in t for all $\pi \in [0, 1]$.

3.3 Optimal Policy Structure

Building on the results of the previous section, we characterize the structure of the optimal policy under condition (4) in Theorem 2.

Theorem 2. *If condition (4) holds, then the optimal policy has one of the following forms:*

(a) For all $t \in \mathbb{N}_0$, there exists a $\pi^*(t) \in [0, 1)$ such that

$$\delta^*(\pi, t) = \begin{cases} C, & 0 \leq \pi \leq \pi^*(t), \\ S, & \pi^*(t) < \pi \leq 1. \end{cases}$$

(b) There exist $t^* \in \mathbb{N}_0$ *and, for all* $t \in \mathbb{N}_0$, $\pi^*(t) \in [0,1)$ *such that, for* $t \leq t^*$,

$$\delta^*(\pi, t) = \begin{cases} C, & 0 \leq \pi \leq \pi^*(t), \\ S, & \pi^*(t) < \pi \leq 1, \end{cases}$$

and for $t > t^*$,

$$\delta^*(\pi, t) = \begin{cases} C, & 0 \leq \pi \leq \pi^*(t), \\ R, & \pi^*(t) < \pi \leq 1, \end{cases}$$

and $\pi^*(t)$ *is nonincreasing in* t *for* $t > t^*$.

Theorem 2 states that the optimal policy has a threshold structure. For all sensor ages $t \in \mathbb{N}_0$, it is optimal to continue operating if the system is known to be in the in-control state ($\pi = 0$), and it is optimal to conduct a full inspection and replace the system if it is known to be in the out-of-control state ($\pi = 1$). Else, if the state of the system is not certain ($\pi \in (0,1)$), to justify the setup cost $c_s > 0$, the probability that the system is in the out-of-control state must exceed a threshold $\pi^*(t)$ to conduct a full inspection. For the decision whether to replace the sensor at a full inspection, there are two possibilities. The first possibility (a) is that the cost c_i of a sensor replacement is prohibitively high and it is never optimal to replace the sensor. The second possibility (b) is that, because a sensor replacement effects a larger reduction in future expected discounted cost if the sensor is older, there is a threshold t^* such that it is optimal to invest c_i and replace the sensor if and only if the sensor age is larger than t^*.

For sensor ages at which it is optimal to replace the sensor at a full inspection, Theorem 2 asserts that the threshold $\pi^*(t)$ is nonincreasing in the sensor age t. The intuition is as follows. Given that a full inspection has the purpose of replacing the sensor as well as eliminating any chance that the system operates in the out-of-control state in the next period, and the benefit of sensor replacement increases with the sensor age, less evidence is needed that the system is in the out-of-control state to conduct a full inspection if the sensor is older. The threshold $\pi^*(t)$ is not, in general, monotone over the range of sensor ages for which it is optimal not to replace the sensor at a full inspection. Without sensor replacement, an older sensor reports noisier observations of the state of the system after both continue operating and full inspection; the only difference is that the probability that the system is in the out-of-control state is zero following a full inspection. The consequence of receiving noisier observations could be that, at a subsequent decision epoch, system operation is continued whereas with more precise observations the decision would have been to conduct a full inspection, or vice versa. This effect causes an increase in future expected discounted cost, but it is unclear whether that increase is higher for $\pi = 0$ or any other $\pi \in [0,1]$; moreover, this may differ from one sensor age to the other. Consequently, $\pi^*(t)$ can be non-monotone in t.

Combining Theorems 1 and 2, we may conclude in general that the optimal policy always has a threshold structure. That is, the optimal policy under condition (2), as established in Theorem 1,

can in principle be viewed as having the threshold structure of Theorem 2 with $\pi^*(t) = 1$ for all $t \in \mathbb{N}_0$.

3.4 Special Case without Sensor Deterioration

A special case arises when the observation matrix does not depend on the sensor age, i.e., $Q(t) = Q$ for all $t \in \mathbb{N}_0$, implying that the sensor does not deteriorate. Note that this case satisfies $Q(t) \succeq_B Q(t+1)$ because $X(t)$ can be taken to be the identity matrix of size $y+1$, for all $t \in \mathbb{N}_0$. It is of interest to study the implications of our structural results in this special case because that allows us to draw a comparison with results available in the literature. These existing results have been developed for maintenance optimization models that, as discussed in the literature review, do not incorporate sensor deterioration.

Without sensor deterioration, there is no point in replacing the sensor, and there is also no reason to adapt decisions to the sensor age. Therefore, in this special case, we may exclude action R from the action space and drop the sensor age t from the state description, so that actions are to be taken solely based on the probability $\pi \in [0,1]$ that the system is in the out-of-control state; we will refer to the model thus obtained as "the simplified model for the special case without sensor deterioration." From Theorems 1 and 2, we have the following corollaries.

Corollary 1. *Consider the simplified model for the special case without sensor deterioration. Let the policy δ be defined by $\delta(\pi) = C$ for all $\pi \in [0,1]$ and the value function V be defined by*

$$V(\pi) = \beta p_{12}(1-\beta)^{-1}(1-\beta(p_{22}-p_{12}))^{-1}c_d + \pi(1-\beta(p_{22}-p_{12}))^{-1}c_d$$

for all $\pi \in [0,1]$. Then, $\delta^* = \delta$ and $V^* = V$ if and only if condition (2) holds.

Corollary 2. *Consider the simplified model for the special case without sensor deterioration. If condition (4) holds, then there exists a $\pi^* \in [0,1)$ such that*

$$\delta^*(\pi) = \begin{cases} C, & 0 \le \pi \le \pi^*, \\ S, & \pi^* < \pi \le 1. \end{cases}$$

Closest to the results of Corollaries 1 and 2, in spite of some differences in the model setup, are results in White [40]. Analogous to our Corollary 1, White's Theorem 5.10(a) identifies a necessary and sufficient condition for the optimality of never taking any maintenance action, and his Lemma 5.9 provides a closed-form expression for the value function attained by this policy. A combination of three of White's results corresponds to our Corollary 2: Lemma 5.8 of White [40] says that the optimal policy has a threshold structure; Lemma 5.7 states that continue operating is the optimal action if it is certain that the system is in the in-control state; and Theorem 5.10(b) states that if the condition of Theorem 5.10(a) is not met, (system) replacement is the optimal action if it is certain that the system is in the out-of-control state.

The model formulated by White [40] is more general than ours in that it has separate actions for inspection and replacement and that it allows for the possibility that observations obtained by costly inspections are imperfect. However, one of the key results for the comparison with Corollaries 1 and 2, Lemma 5.8 in which the threshold structure is established, is derived under the condition that observations obtained by a costly inspection are less informative than costless observations, which ensures that it is never optimal to conduct an inspection. Without inspection (because it is also assumed that the cost of a replacement does not depend on the state of the system) the model of White [40] reduces to a model similar to ours with $c_r = 0$ (as in Remark 2). That our results show the optimality of a threshold policy when $c_r > 0$ is important because in practice a maintenance engineer often first inspects the system to assess the degree of maintenance needed before taking any actions, and then the cost of maintenance depends on the state of the system. There is another respect in which Corollaries 1 and 2 are more general than the corresponding results of White [40]. Whereas White [40] assumes that $p_{22} = 1$, we make no assumptions on the transition probability matrix. Because most systems cannot recover from an out-of-control state by themselves, this generalization is mainly of theoretical interest.

Other papers provide a less complete characterization of the optimal policy structure or make at least as strong assumptions on the transition probability matrix and the cost parameters. For example, Givon and Grosfeld-Nir [12] develop a model for the replacement of TV shows that can also be applied to the replacement of binary-state systems, and their Proposition 1 is akin to our Corollary 1, but they assume that $c_r = 0$ and $p_{22} = 1$. Structural results by Ohnishi et al. [30] on the optimal inspection and replacement policy for multi-state systems imply a threshold structure for the optimal policy in the binary-state case, but their results do not characterize this threshold structure in the same detail as in Corollaries 1 and 2, and their assumptions require $c_d \geq c_r$ and $p_{22} \geq p_{12}$. The special case we consider here also is a special case of the model of Dada and Marcellus [7], who distinguish between "routine maintenance" and "learning maintenance." When the cost of learning maintenance is so high that such maintenance is never optimal, their Proposition 2 is similar to our Corollary 1, and their Proposition 7 is similar to our Corollary 2. Their assumptions include $c_r = 0$ and $p_{22} = 1$.

4 Value of Coordinating System and Sensor Maintenance

Theorem 2 of Section 3.3 shows that the optimal policy coordinates system and sensor maintenance, if it is ever optimal to replace the sensor. On the one hand, when the sensor age exceeds t^*, and it is optimal to replace the sensor at the earliest opportunity, a full inspection is performed only if the probability that the system is out of control justifies doing so. On the other hand, the threshold for performing a full inspection, $\pi^*(t)$, is adapted to the sensor age t in that it is lower when the benefit of sensor replacement is higher, i.e., $\pi^*(t)$ decreases in t for $t > t^*$. To assess

the value of coordinating system and sensor maintenance, in this section we present numerical examples to benchmark the optimal policy against two heuristic policies that apply a weaker form of coordination.

4.1 Heuristics

The first heuristic policy, termed heuristic H_1, relaxes the coordination between system and sensor maintenance by using a sensor-age-independent threshold on π. Specifically, heuristic H_1 is defined as the best policy (yielding the lowest total expected discounted cost for initial information state $(0,0)$) within the class of policies that can be described by a threshold $\pi^{H_1} \in [0,1]$ as

$$\delta(\pi,t) = \begin{cases} C, & \pi \leq \pi^{H_1}, \\ S, & \pi > \pi^{H_1}, \end{cases}$$

or by thresholds $\pi^{H_1} \in [0,1]$ and $t^{H_1} \in \mathbb{N}_0$ as

$$\delta(\pi,t) = \begin{cases} C, & \pi \leq \pi^{H_1}, \\ S, & \pi > \pi^{H_1},\ t \leq t^{H_1}, \\ R, & \pi > \pi^{H_1},\ t > t^{H_1}. \end{cases} \tag{5}$$

Thus, heuristic H_1 prescribes a full inspection whenever $\pi > \pi^{H_1}$, and the sensor is either never replaced, or replaced at every full inspection when the sensor age exceeds a threshold t^{H_1}. This heuristic is system-directed in the sense that it is the probability that the system is in the out-of-control state that determines whether or not a full inspection is conducted, and the threshold π^{H_1} is not adapted to the sensor age. We denote its total expected discounted cost, as a function of the initial information state $(\pi,t) \in \Omega$, by $V^{H_1}(\pi,t)$.

The second heuristic policy, termed heuristic H_2, conducts a full inspection if either π exceeds a sensor-age-independent threshold, or t exceeds an age threshold, or both. Specifically, heuristic H_2 is the best policy within the class of policies described by a threshold $\pi^{H_2} \in [0,1]$ as

$$\delta(\pi,t) = \begin{cases} C, & \pi \leq \pi^{H_2}, \\ S, & \pi > \pi^{H_2}, \end{cases}$$

or by thresholds $\pi^{H_2} \in [0,1]$ and $t^{H_2} \in \mathbb{N}_0$ as

$$\delta(\pi,t) = \begin{cases} C, & \pi \leq \pi^{H_2},\ t \leq t^{H_2}, \\ S, & \pi > \pi^{H_2},\ t \leq t^{H_2}, \\ R, & t > t^{H_2}. \end{cases}$$

Under this heuristic, the sensor is replaced periodically because either it is never replaced (i.e., the interval between sensor replacements is infinite), or a full inspection with sensor replacement is

performed every $t^{H_2}+1$ periods. A full inspection without sensor replacement is performed when $\pi > \pi^{H_2}$ if the sensor age is t^{H_2} or lower, or if the sensor is never replaced. This heuristic employs the least coordination between system and sensor maintenance: the only coordination is when the sensor age prompts a full inspection to replace the sensor, the system is also replaced if it is found to be in the out-of-control state. We denote its total expected discounted cost, as a function of the initial information state $(\pi, t) \in \Omega$, by $V^{H_2}(\pi, t)$.

We note that in exchange for sacrificing coordination between system and sensor maintenance, the heuristic policies have a simpler structure, leading to easier implementation. In the heuristic H_1, the milder form of coordination eliminates the need to store a different threshold value for each sensor age. Heuristic H_2 has the additional advantage that some full inspections can be planned in advance.

4.2 Solution Method

If condition (2) is satisfied, then Theorem 1 provides a closed-form expression for the optimal value function and indicates that it is optimal to never apply any maintenance. Since the heuristic policies achieve the same behavior by setting $\pi^{H_1} = \pi^{H_2} = 1$, the optimal policy and the heuristics coincide. However, here we will be interested in settings in which condition (4) is satisfied. Therefore, we need methods to compute the optimal and heuristic policies numerically.

Our numerical examples consider instances in which, at some sensor age, the sensor stops deteriorating and reaches a stationary level of informativeness. That is, there exists a minimum sensor age t_{max} such that $Q(t) = Q(t_{max})$ for all $t \geq t_{max}$. As an example of the stationary level of informativeness that may be reached, sensor observations are completely uninformative beyond t_{max} if the observation matrix $Q(t)$, $t \geq t_{max}$, has identical rows, i.e., $q_{1k}(t) = q_{2k}(t)$ for all $k \in \mathcal{O}$ (it then holds that $\widehat{Q} \succeq_B Q(t)$ for all observation matrices $\widehat{Q}$). Given that sensor observations are equally informative for all sensor ages $t \geq t_{max}$, in our numerical examples, we truncate the countably infinite state space $\mathcal{S} \times \mathbb{N}_0$ to the finite state space $\mathcal{S} \times \{0, 1, \ldots, t_{max}\}$, in which we reinterpret the sensor age in states (i, t_{max}), $i \in \mathcal{S}$, as being larger than or equal to t_{max}. Consequently, the information state space is denoted by $\widetilde{\Omega} = [0, 1] \times \{0, 1, \ldots, t_{max}\}$.

Unfortunately, infinite-horizon POMDP models are computationally intractable, even with a finite underlying state space. Although methods have been developed to compute ϵ-optimal policies (see overviews by Lovejoy [24] and Poupart [31]), these have limited applicability, as the computational burden grows exponentially with the cardinality of the observation space, as well as the state and action spaces. Grid-based approximation techniques are a widely used alternative (Lovejoy [24], Hauskrecht [16]). The basic idea is to approximate the optimal value function using interpolation between a finite number of grid points in the information state space, and then to derive an approximate optimal policy. It is this approach that we will use to solve the numerical examples.

One reason why it is especially attractive to use grid-based techniques for our model is that we have two possible system states. Generally, in a uniform grid, an increase in grid resolution (i.e., the number of intervals in which each dimension of the belief space is subdivided) leads to an exponential growth of the number of grid points (Hauskrecht [16]); however, in our case, the belief about the state of the system is expressed as a scalar, and the number of grid points and the associated computational requirements grow only linearly in the resolution. Another advantage is that by choosing different interpolation rules, we can utilize the structural results of Section 3 to generate bounds on the optimal value function. Finally, we can easily evaluate a given policy using a grid-based approximation, which is useful for constructing and evaluating the heuristic policies (since that requires identification of the best policy within a class of policies); for evaluating a given policy no exact method is known to exist (Hansen [15]).

To facilitate the exposition of the solution method described below, we define the function H by $H(\pi, t, V) = \min_{a \in \mathcal{A}} H^a(\pi, t, V)$ for all $(\pi, t) \in \tilde{\Omega}$ and $V \in \mathcal{V}$, so that we can succinctly write the optimality equations in (1) as $V(\pi, t) = H(\pi, t, V)$ for all $(\pi, t) \in \tilde{\Omega}$. Our grid-based approach is then detailed as follows. Initially, we set a grid resolution $z \in \mathbb{N}$ and define a set of $z + 1$ equally spaced points in the interval $[0, 1]$, $\mathcal{G} = \{0, 1/z, \ldots, 1\}$, to construct the uniform grid $\bar{\Omega} = \mathcal{G} \times \{0, 1, \ldots, t_{max}\}$. For deriving a lower bound on the optimal value function, following Lovejoy [25], we replace the function H in the optimality equations with another function H_L that performs a dynamic programming update only at grid points in $\bar{\Omega}$ and applies linear interpolation in between. That is, for all $V \in \mathcal{V}$, H_L is defined by $H_L(\pi, t, V) = H(\pi, t, V)$ for $(\pi, t) \in \bar{\Omega}$ and

$$H_L(\pi, t, V) = (\lceil \pi z \rceil - \pi z) H(\lfloor \pi z \rfloor / z, t, V) + (\pi z - \lfloor \pi z \rfloor) H(\lceil \pi z \rceil / z, t, V)$$

for $(\pi, t) \in \widetilde{\Omega} \setminus \bar{\Omega}$, where $\lceil \pi z \rceil$ denotes the smallest integer larger than or equal to πz, and $\lfloor \pi z \rfloor$ denotes the largest integer smaller than or equal to πz. The set of equations $V(\pi, t) = H_L(\pi, t, V)$ for all $(\pi, t) \in \widetilde{\Omega}$ are the optimality equations associated with an MDP in which updating the belief about the state of the system is done by stochastically rounding the output of Bayes' rule to one of the nearest elements in $\mathcal{G}$. We use a non-convex interpolation rule (in the terminology of Hauskrecht [16]) to derive an upper bound. For all $V \in \mathcal{V}$, the function H_U is defined by $H_U(\pi, t, V) = H(\pi, t, V)$ for $(\pi, t) \in \bar{\Omega}$ and

$$H_U(\pi, t, V) = H(\lceil \pi z \rceil / z, t, V) - (\lceil \pi z \rceil / z - \pi) c_r$$

for $(\pi, t) \in \widetilde{\Omega} \setminus \bar{\Omega}$. The set of equations $V(\pi, t) = H_U(\pi, t, V)$ for all $(\pi, t) \in \widetilde{\Omega}$ are the optimality equations associated with an MDP in which the output of Bayes' rule is rounded up and the immediate expected cost of all actions is lowered. The following theorem states the relationship between the solutions to the optimality equations obtained using H_L, H, and H_U.

Theorem 3. *Let V_L and V_U be value functions such that $V_L(\pi, t) = H_L(\pi, t, V_L)$ and $V_U(\pi, t) =$*

$H_U(\pi,t,V_U)$ *for all* $(\pi,t)\in\widetilde{\Omega}$. *If condition* (4) *holds, then* $V_L(\pi,t)\le V^*(\pi,t)\le V_U(\pi,t)$ *for all* $(\pi,t)\in\widetilde{\Omega}$.

The proof of Theorem 3, which is in the Appendix, uses general principles for POMDPs to establish the validity of the lower bound. That V_U provides an upper bound is specific to our problem and relies on the analytical result on the form of the optimal value function obtained in Lemma 1.

We use value iteration algorithms based on the functions H_L and H_U to compute the value functions V_L and V_U as defined in Theorem 3. These algorithms converge to V_L and V_U and require computing values of the iterates $V_{L,n}$ and $V_{U,n}$, in all iterations $n\in\mathbb{N}$, only at grid points in $\bar{\Omega}$ (cf. Hauskrecht [16]). Upon terminating with bounds on the optimal value function, we construct an approximate optimal policy $\bar{\delta}^*$ using the lower bound. We let $\bar{\delta}^*(\pi,t)=\arg\min_{a\in\mathcal{A}} H_L^a(\pi,t,V_L)$ for all $(\pi,t)\in\widetilde{\Omega}$, where we define, for all $a\in\mathcal{A}$ and $V\in\mathcal{V}$, $H_L^a(\pi,t,V)=H^a(\pi,t,V)$ for $(\pi,t)\in\bar{\Omega}$ and

$$H_L^a(\pi,t,V)=(\lceil\pi z\rceil-\pi z)H^a(\lfloor\pi z\rfloor/z,t,V)+(\pi z-\lfloor\pi z\rfloor)H^a(\lceil\pi z\rceil/z,t,V)$$

for $(\pi,t)\in\widetilde{\Omega}\setminus\bar{\Omega}$.

To obtain an approximation for heuristic H_1, we enumerate all policies determined by the threshold values $(\pi^{H_1},t^{H_1})\in\bar{\Omega}$ via (5). (This search space includes policies in which the sensor is never replaced — with the truncated state space, this behavior is achieved by setting $t^{H_1}=t_{max}$.) For each policy δ, we approximate its total expected discounted cost by computing the value function $\bar{V}^\delta$ that satisfies $\bar{V}^\delta(\pi,t)=H_L^{\delta(\pi,t)}(\pi,t,\bar{V}^\delta)$ for all $(\pi,t)\in\widetilde{\Omega}$. Thus, we solve a system of $|\bar{\Omega}|$ linear equations to obtain $\bar{V}^\delta(\pi,t)$ for all $(\pi,t)\in\bar{\Omega}$ and use linear interpolation to determine $\bar{V}^\delta(\pi,t)$ for all $(\pi,t)\in\widetilde{\Omega}\setminus\bar{\Omega}$. We take as the approximate H_1 heuristic the policy δ that yields the lowest value of $\bar{V}^\delta(0,0)$ and let $\bar{V}^{H_1}=\bar{V}^\delta$ be the corresponding approximate total expected discounted cost. An analogous procedure is used to obtain the approximate H_2 heuristic and its approximate total expected discounted cost $\bar{V}^{H_2}$.

4.3 Numerical Examples

Considered are two examples, which both use the discount factor $\beta=0.999$. As announced in Section 4.2, the other problem parameters are chosen such that condition (4) is satisfied; therefore, the optimal policies have the structure described in Theorem 2. We compute the (approximate) optimal policy using $z=5000$ and the (approximate) heuristic policy H_1 and heuristic policy H_2 using $z=500$ as the grid resolution. Optimality gaps of the heuristic policies are calculated via $(\bar{V}^{H_1}(0,0)-V_L(0,0))/V_L(0,0)$ and $(\bar{V}^{H_2}(0,0)-V_L(0,0))/V_L(0,0)$.

Example 1. Let the cost parameters be $c_d=100$, $c_s=75$, $c_r=50$, and $c_i=20$, and let the state

of the system evolve according to the transition probability matrix

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix}.$$

Suppose that the sensor observation is binomially distributed with parameters y and $p_i(t)$, where $p_i(t)$ depends on the state of the system $i \in \mathcal{S}$ and the sensor age $t \in \mathbb{N}_0$. Thus, for all $t \in \mathbb{N}_0$, the elements of observation matrix $Q(t)$ are given by

$$q_{ik}(t) = \binom{y}{k} p_i(t)^k (1 - p_i(t))^{y-k},$$

for all $i \in \mathcal{S}$, $k \in \mathcal{O}$. Let $y = 50$, and let $p_1(t) = 0.4 + 0.015t$ and $p_2(t) = 0.7 - 0.015t$ for $t \leq 10$, and $p_1(t) = p_2(t) = 0.55$ for $t > 10$. This implies that the sensor stops deteriorating at age $t_{max} = 10$, and when the sensor's age is higher than 10, its observations are completely uninformative.

[Figure 1 about here.]

With $p_1(t) \leq p_2(t)$ for all $t \in \mathbb{N}_0$, if $p_1(t)$ is increasing in t and $p_2(t)$ is decreasing in t, as in this example, then sensor observations become less discriminatory as the sensor ages; Figure 1 illustrates how the conditional probability distributions of the sensor observation, given the state of the system, converge as t increases to t_{max}. Intuitively, this means that the performance of the sensor decreases over time, i.e., the sensor is subject to deterioration. Indeed, it can be established that the assumption $Q(t) \succeq_B Q(t+1)$, for all $t \in \mathbb{N}_0$, is satisfied. One can verify that for $t < t_{max}$ the matrix $X(t)$ with elements

$$x_{kl}(t) = \sum_{w=\max\{l-k,0\}}^{\min\{l,y-k\}} \binom{k}{l-w} \xi(t)^{l-w} (1 - \xi(t))^{k-l+w} \binom{y-k}{w} \zeta(t)^w (1 - \zeta(t))^{y-k-w} \tag{6}$$

for all $k, l \in \mathcal{O}$, where

$$\xi(t) = \frac{p_2(t+1) - p_1(t+1) + p_1(t+1)p_2(t) - p_1(t)p_2(t+1)}{p_2(t) - p_1(t)},$$
$$\zeta(t) = \frac{p_1(t+1)p_2(t) - p_1(t)p_2(t+1)}{p_2(t) - p_1(t)},$$

is a stochastic matrix such that $Q(t+1) = Q(t)X(t)$. For $t \geq t_{max}$, $X(t)$ can simply be taken to be the identity matrix of dimension $y + 1$.

The elements of the matrix $X(t)$ defined via (6) can be given a probabilistic interpretation. The idea is that an observation by a sensor of age $t + 1$, which may be seen as the outcome of a binomial experiment with y trials and success probability $p_i(t + 1)$, may also be viewed as the outcome of a two-stage experiment. The first stage is a binomial experiment with y trials and success probability $p_i(t)$, whose outcome corresponds to an observation by a sensor of age t, and the second stage adds random noise. Specifically, in the second stage, a success trial is changed to a

failure with probability $1-\xi(t)$ and unchanged with probability $\xi(t)$, and a failure trial is changed to a success with probability $\zeta(t)$ and unchanged with probability $1-\zeta(t)$. (That $\xi(t)$ and $\zeta(t)$ are in $[0,1]$ follows from our assumptions on $p_1(t)$ and $p_2(t)$.) The equivalence holds because $\xi(t)$ and $\zeta(t)$ are such that $p_i(t)\xi(t)+(1-p_i(t))\zeta(t)=p_i(t+1)$ for all $i\in\mathcal{S}$. It is important to note that, conditionally on the number of successes in the first stage, the outcome of the two-stage experiment is independent of the state of the system $i\in\mathcal{S}$. Therefore, the relationship $Q(t+1)=Q(t)X(t)$ can be satisfied by letting $x_{kl}(t)$, for all $k,l\in\mathcal{O}$, be the conditional probability that, given that the first stage resulted in k successes, the outcome of the experiment is l. Since the experiment's outcome is distributed as the sum of two independent binomial random variables, one with parameters k and $\xi(t)$ and one with parameters $y-k$ and $\zeta(t)$, application of the discrete convolution formula yields Equation (6).

[Figures 2–4 about here.]

The optimal policy, the H_1 heuristic policy, and the H_2 heuristic policy for this example are depicted in Figures 2–4. It can be seen that in the optimal policy, when a full inspection is conducted, the sensor is replaced if its age is higher than $\bar{t}^*=2$. Furthermore, in accordance with Theorem 2, the threshold $\bar{\pi}^*(t)$ is nonincreasing in t for $t>\bar{t}^*$; it is lowest when the sensor age is $t_{max}=10$ or higher. The policy structure also exemplifies that, as we discussed in Section 3.3, the threshold $\bar{\pi}^*(t)$ may not be monotone in t for $t\leq\bar{t}^*$. Looking at the heuristic policies, we see that heuristic H_1 uses similar threshold values as the optimal policy. The sensor age threshold is the same ($\bar{t}^{H_1}=\bar{t}^*$), and the threshold on the probability that the system is in the out-of-control state is between the lowest and highest value of the optimal sensor-age-dependent threshold ($\min_t\bar{\pi}^*(t)<\bar{\pi}^{H_1}<\max_t\bar{\pi}^*(t)$). The H_2 heuristic uses much higher threshold values ($\bar{t}^{H_2}>\bar{t}^*$, while $\bar{\pi}^{H_2}$ is close to $\max_t\bar{\pi}^*(t)$). The differences in performance are in line with these policy differences: for initial information state $(0,0)$, the bounds on the total expected discounted cost under the optimal policy are $V_L(0,0)=23{,}931.7$ and $V_U(0,0)=23{,}946.8$, and the heuristic policies achieve total expected discounted costs $\bar{V}^{H_1}(0,0)=24{,}175.1$ and $\bar{V}^{H_2}(0,0)=26{,}781.9$. Thus, whereas the optimality gap of the H_1 heuristic is only 1.0%, using heuristic H_2 results in an increase in total expected discounted cost of 11.9% relative to the optimal policy.

The policy differences between the H_2 heuristic and the optimal policy observed in Example 1, as well as the corresponding difference in performance, can be explained as follows. Using this heuristic policy, when the sensor age exceeds t^{H_2}, a full inspection with sensor replacement is always performed. However, if there is no indication that the system is in the out-of-control state, conducting a full inspection is inefficient, and it is more beneficial to continue operation and postpone full inspection. Therefore, to diminish the chance that the probability that the system is in the out-of-control state is very small when the sensor age reaches $t^{H_2}+1$, a relatively high value is used for t^{H_2}. Consequently, for some sensor ages less than $t^{H_2}+1$, the sensor is not replaced at

a full inspection while, given that a full inspection is conducted, it would have been beneficial to do so. Hence full inspections are less attractive, so the value for π^{H_2} is also relatively high.

Realistically, for a safety-critical system, the optimal thresholds on the probability that the system is in the out-of-control state might be considerably lower than in Example 1. Therefore, in the next example, we choose parameter values for which the thresholds are lower. It turns out that the comparison between the policies remains virtually the same.

Example 2. Let the cost parameters be $c_d = 500$, $c_s = 50$, $c_r = 100$, and $c_i = 10$, and let the state of the system evolve according to the transition probability matrix

$$P = \begin{pmatrix} 0.98 & 0.02 \\ 0 & 1 \end{pmatrix}.$$

Suppose that the sensor observation is binomially distributed with parameters Y and p_i, where Y is itself a binomially distributed random variable with parameters y and $\theta(t)$, p_i depends on the state of the system $i \in \mathcal{S}$, and $\theta(t)$ depends on the sensor age $t \in \mathbb{N}_0$. This compound distribution is in fact equivalent to a binomial distribution with parameters y and $p_i\theta(t)$, so for all $t \in \mathbb{N}_0$, the elements of observation matrix $Q(t)$ are given by

$$q_{ik}(t) = \binom{y}{k} (p_i\theta(t))^k (1 - p_i\theta(t))^{y-k},$$

for all $i \in \mathcal{S}$, $k \in \mathcal{O}$. Let $y = 50$, let $p_1 = 0.3$ and $p_2 = 0.6$, and let $\theta(t) = 1 - 0.1t$ for $t \leq 10$, and $\theta(t) = 0$ for $t > 10$. Thus, again, sensor deterioration ends at sensor age $t_{max} = 10$, and observations from an older sensor are completely uninformative.

[Figure 5 about here.]

If $\theta(t)$ is decreasing in t, as in this example, then the sensor generates a weaker signal as it ages (the random variable Y is stochastically smaller for a higher sensor age t), resulting in lower observations. Figure 5 depicts how, for both states of the system, the sensor observation gets stochastically smaller over time until it equals zero with probability 1. That sensor observations carry less information with a weaker signal is intuitive. It can be established that the assumption that, for all $t \in \mathbb{N}_0$, $Q(t) \succeq_B Q(t+1)$ is satisfied by verifying that, for $t < t_{max}$, the matrix $X(t)$ with elements

$$x_{kl}(t) = \begin{cases} \binom{k}{l} \left(\frac{\theta(t+1)}{\theta(t)}\right)^l \left(1 - \frac{\theta(t+1)}{\theta(t)}\right)^{k-l}, & k \geq l, \\ 0, & k < l, \end{cases}$$

is a stochastic matrix such that $Q(t+1) = Q(t)X(t)$. This matrix corresponds to a similar noise corruption of sensor observations as was described in Example 1, but now with parameters $\xi(t) = \theta(t+1)/\theta(t)$ and $\zeta(t) = 0$.

[Figures 6–8 about here.]

Figures 6–8 respectively depict the optimal policy, the H_1 heuristic policy, and the H_2 heuristic policy. The H_2 policy deviates more from the optimal policy and uses higher threshold values than the H_1 policy. The total expected discounted cost is bounded by $V_L(0,0) = 11{,}507.2$ and $V_U(0,0) = 11{,}574.6$ for the optimal policy, and is $\bar{V}^{H_1}(0,0) = 11{,}684.1$ and $\bar{V}^{H_2}(0,0) = 13{,}208.5$ for the heuristics, showing that the optimality gap of the H_2 heuristic (14.8%) is significantly larger than that of the H_1 heuristic (1.5%).

5 Conclusions

In this paper, we have studied the joint optimal system and sensor maintenance policy for a safety-critical system monitored by a deteriorating sensor. We developed a POMDP model to address the problem of scheduling full inspections and sensor replacements to minimize the infinite-horizon total expected discounted cost. In it, sensor deterioration is modeled by means of the Blackwell order. We derived a necessary and sufficient condition for the optimality of never applying any maintenance, and we showed that, in general, the optimal policy has a threshold structure with respect to the probability that the system is in the out-of-control state and the sensor age. These structural results are theoretically relevant also for the special case without sensor deterioration. Additionally, we provided numerical examples to highlight how the total expected discounted cost increases with a lower degree of coordination between system and sensor maintenance.

In the setting we studied, only one sensor is used to monitor the state of a safety-critical system. However, many condition monitoring systems combine multiple sensors to measure the state of a system. Therefore, a natural direction for future research is to examine an extension of our model in which the safety-critical system can be monitored by multiple deteriorating sensors, or a combination of deteriorating and non-deteriorating sensors. Computation of the (approximate) optimal policy may become difficult because the age of all deteriorating sensors needs to be considered in deciding on maintenance actions, and as such, most dynamic programming methods will suffer from the curse of dimensionality. Yet, it might be possible to achieve an acceptable performance with approximate dynamic programming methods. More importantly, we expect that sensor maintenance and coordination between system and sensor maintenance will still be important elements of the optimal maintenance policy, and that such insights could help to guide the development of effective heuristic policies.

It would also be interesting to consider alternate forms of sensor deterioration. We assumed that the sensor observations only depend on the state of the system, according to a probabilistic relation determined by the sensor age. Although this assumption is convenient because it allows decisions to be based on the probability that the system is in the out-of-control state and the (known) sensor age, in reality sensor observations might also depend on previous sensor observations, and the relation between sensor observations and the state of the system might change randomly as

the sensor ages. Finally, whereas our focus was on binary-state systems, future research could investigate the optimal maintenance policy when the condition of the system can be classified into multiple states. At a full inspection, it will then need to be decided whether or not to replace the system and the sensor depending on the state in which the system is found.

Acknowledgements

The authors thank the Associate Editor and two anonymous referees for their constructive comments and suggestions. The first author was supported in part by the Prins Bernhard Cultuurfonds.

References

[1] S.C. Albright. Structural results for partially observable Markov decision processes. *Operations Research*, 27(5):1041–1053, 1979.

[2] S. Alizamir, F. de Véricourt, and P. Sun. Diagnostic accuracy under congestion. *Management Science*, 59(1):157–171, 2013.

[3] J.L. Blackshire, V. Giurgiutiu, A. Cooney, and J. Doane. Characterization of sensor performance and durability for structural health monitoring systems. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 5770:66–74, 2005.

[4] D. Blackwell. Comparison of experiments. In *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, 93–102, 1951.

[5] D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

[6] J. Coble, P. Ramuhalli, R. Meyer, H. Hashemian, B. Shumaker, and D. Cummins. Calibration monitoring for sensor calibration interval extension: identifying technical gaps. In *Proceedings of the Future of Instrumentation International Workshop*, 2012.

[7] M. Dada and R. Marcellus. Process control with learning. *Operations Research*, 42(2):323–336, 1994.

[8] A. Davies. *Handbook of Condition Monitoring: Techniques and Methodology*. Chapman & Hall, 1998.

[9] K. Fowler, editor. *Mission-Critical and Safety-Critical Systems Handbook: Design and Development for Embedded Applications*. Newnes, 2009.

[10] A. Ghasemi, S. Yacout, and M.S. Ouali. Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research*, 45(4):989–1012, 2007.

[11] S.M. Gilbert and H.M. Bar. The value of observing the condition of a deteriorating machine. *Naval Research Logistics*, 46(7):790–808, 1999.

[12] M. Givon and A. Grosfeld-Nir. Using partially observable Markov processes to select optimal termination time of TV shows. *Omega*, 36(3):477–485, 2008.

[13] A. Grosfeld-Nir. Control limits for two-state partially observable Markov decision processes. *European Journal of Operational Research*, 182(1):300–304, 2007.

[14] W.R. Habel and A. Bismarck. Optimization of the adhesion of fiber-optic strain sensors embedded in cement matrices; a study into long-term fiber strength. *Journal of Structural Control*, 7(1):51–76, 2000.

[15] E.A. Hansen. An improved policy iteration algorithm for partially observable MDPs. In *Proceedings of the 10th Annual Conference on Advances in Neural Information Processing Systems (NIPS)*, 1015–1021, 1997.

[16] M. Hauskrecht. Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 13:33–94, 2000.

[17] S. Karlin. *Total Positivity*. Stanford University Press, 1968.

[18] M.J. Kim and V. Makis. Joint optimization of sampling and control of partially observable failing systems. *Operations Research*, 61(3):777–790, 2013.

[19] V. Krishnamurthy and D.V. Djonin. Structured threshold policies for dynamic sensor scheduling—a partially observed Markov decision process approach. *IEEE Transactions on Signal Processing*, 55(10):4938–4957, 2007.

[20] Y. Kuo. Optimal adaptive control policy for joint machine maintenance and product quality control. *European Journal of Operational Research*, 171(2):586–597, 2006.

[21] L. Le Cam. Comparison of experiments: a short review. In T. Ferguson and L. Shapley, editors, *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, Lecture Notes–Monograph Series, 127–138. IMS, 1996.

[22] M. Lévesque and L.M. Maillart. Business opportunity assessment with costly, imperfect information. *IEEE Transactions on Engineering Management*, 55(2):279–291, 2008.

[23] W.S. Lovejoy. Some monotonicity results for partially observed Markov decision processes. *Operations Research*, 35(5):736–743, 1987.

[24] W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28(1):47–66, 1991.

[25] W.S. Lovejoy. Computationally feasible bounds for partially observed Markov decision processes. *Operations Research*, 39(1):162–175, 1991.

[26] L.M. Maillart and L. Zheltova. Structured maintenance policies on interior sample paths. *Naval Research Logistics*, 54(6):645–655, 2007.

[27] L.M. Maillart, T.G. Yeung, and Z.G. Icten. Selecting test sensitivity and specificity parameters to optimally maintain a degrading system. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 225(2):131–139, 2011.

[28] G.E. Monahan. A survey of partially observable Markov decision processes: theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

[29] M. Naghshvar and T. Javidi. Information utility in active sequential hypothesis testing. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*,

123–129, 2010.

[30] M. Ohnishi, H. Kawai, and H. Mine. An optimal inspection and replacement policy under incomplete state information. *European Journal of Operational Research*, 27(1):117–128, 1986.

[31] P. Poupart. *Exploiting structure to efficiently solve large scale partially observable Markov decision processes*. PhD thesis, University of Toronto, 2005.

[32] M.L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 2005.

[33] A. Ray and S. Phoha. Calibration and estimation of redundant signals for real-time monitoring and control. *Signal Processing*, 83(12):2593–2605, 2003.

[34] U. Rieder. Structural results for partially observed control models. *Zeitschrift für Operations Research*, 35(6):473–490, 1991.

[35] D. Rosenfield. Markovian deterioration with uncertain information. *Operations Research*, 24 (1):141–155, 1976.

[36] S.M. Ross. Quality control under Markovian deterioration. *Management Science*, 17(2):587–596, 1971.

[37] R.D. Smallwood and E.J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5):1071–1088, 1973.

[38] R. Srinivasan and A.K. Parlikad. Value of condition monitoring in infrastructure maintenance. *Computers & Industrial Engineering*, 66(2):233–241, 2013.

[39] C. Ulu and J.E. Smith. Uncertainty, information acquisition, and technology adoption. *Operations Research*, 57(3):740–752, 2009.

[40] C.C. White, III. A Markov quality control process subject to partial observation. *Management Science*, 23(8):843–852, 1977.

[41] C.C. White, III. Optimal control-limit strategies for a partially observed replacement problem. *International Journal of Systems Science*, 10(3):321–331, 1979.

[42] C.C. White, III and D.P. Harrington. Application of Jensen's inequality to adaptive suboptimal design. *Journal of Optimization Theory and Applications*, 32(1):89–99, 1980.

[43] H. Zhang. Partially observable Markov decision processes: a geometric technique and analysis. *Operations Research*, 58(1):214–228, 2010.

Appendix

Proof of Theorem 1. For notational brevity, define $d \equiv \beta p_{12}(1-\beta)^{-1}(1-\beta(p_{22}-p_{12}))^{-1}c_d$ and $b \equiv (1-\beta(p_{22}-p_{12}))^{-1}c_d$. Then, the value function in (3) can be expressed as $V(\pi,t) = d + \pi b$ for all $(\pi,t) \in \Omega$ and condition (2) can be written as $b \leq c_s + c_r$. It is easy to verify that $d = \beta(d + p_{12}b)$ and $b = c_d + \beta(p_{22}-p_{12})b$. Therefore, for all $(\pi,t) \in \Omega$,

$$\begin{aligned}
H^C(\pi,t,V) &= \pi c_d + \beta \sum_{k\in\mathcal{O}} \sigma(k;\pi,t) V(\psi(\pi,t,k),t+1) \\
&= \pi c_d + \beta \sum_{k\in\mathcal{O}} \sigma(k;\pi,t)(d + \psi(\pi,t,k)b) \\
&= \pi c_d + \beta d + \beta \sum_{k\in\mathcal{O}} ((1-\pi)p_{12} + \pi p_{22}) q_{2k}(t+1) b \\
&= \pi c_d + \beta d + \beta p_{12} b + \pi\beta(p_{22}-p_{12})b = d + \pi b; \\
H^S(\pi,t,V) &= H^C(0,t,V) + c_s + \pi c_r = d + c_s + \pi c_r; \\
H^R(\pi,t,V) &= H^C(0,0,V) + c_s + c_i + \pi c_r = d + c_s + c_i + \pi c_r.
\end{aligned}$$

Note that $V(\pi,t) = H^C(\pi,t,V)$ for all $(\pi,t) \in \Omega$; thus, V gives the total expected discounted cost attained by policy δ. Next, we check whether V satisfies the optimality equations. If condition (2) holds, then

$$\min_{a\in\mathcal{A}} H^a(\pi,t,V) = H^C(\pi,t,V) = V(\pi,t)$$

for all $(\pi,t) \in \Omega$, so we conclude that $\delta = \delta^*$ and $V = V^*$. If condition (2) does not hold, then

$$\min_{a\in\mathcal{A}} H^a(1,t,V) = H^S(1,t,V) < V(1,t)$$

for all $t \in \mathbb{N}_0$, so we conclude that $\delta \neq \delta^*$ and $V > V^*$, meaning $V(\pi,t) \geq V^*(\pi,t)$ for all $(\pi,t) \in \Omega$ with strict inequality for some $(\pi,t) \in \Omega$ (in particular, $V(1,t) > V^*(1,t)$ for all $t \in \mathbb{N}_0$). □

The proof of Lemma 1 will employ the monotonic relationships established in the following proposition.

Proposition 1. *Let $t \in \mathbb{N}_0$. Suppose $Q(t+1)$ is TP_2 and such that, for all $k,l \in \mathcal{O}$ with $k \leq l$, $q_{2k}(t+1) > 0$ implies $q_{2l}(t+1) > 0$.*

(i) Let $\pi \in [0,1]$ and $k,l \in \mathcal{O}$ such that $k \leq l$. Then $\psi(\pi,t,k) \leq \psi(\pi,t,l)$.

(ii) Let $\pi,\hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $k \in \mathcal{O}$. If $p_{22} \geq (\leq)\ p_{12}$, then $\psi(\pi,t,k) \leq (\geq)\ \psi(\hat{\pi},t,k)$.

(iii) Let $\pi,\hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $l \in \mathcal{O}$. If $p_{22} \geq (\leq)\ p_{12}$, then $\sum_{k\geq l}\sigma(k;\pi,t) \leq (\geq)\ \sum_{k\geq l}\sigma(k;\hat{\pi},t)$.

Proof of Proposition 1.

(i) Clearly, if $\psi(\pi,t,k)=0$, then $\psi(\pi,t,k)\leq\psi(\pi,t,l)$. Else if $\psi(\pi,t,k)>0$, it must be that $q_{2k}(t+1)>0$ and, therefore, $q_{2l}(t+1)>0$. Consequently, the TP_2 property of $Q(t+1)$ can be used to obtain

$$\begin{aligned}\frac{1-\psi(\pi,t,k)}{\psi(\pi,t,k)}&=\left(\frac{(1-\pi)p_{11}+\pi p_{21}}{(1-\pi)p_{12}+\pi p_{22}}\right)\left(\frac{q_{1k}(t+1)}{q_{2k}(t+1)}\right)\\&\geq\left(\frac{(1-\pi)p_{11}+\pi p_{21}}{(1-\pi)p_{12}+\pi p_{22}}\right)\left(\frac{q_{1l}(t+1)}{q_{2l}(t+1)}\right)=\frac{1-\psi(\pi,t,l)}{\psi(\pi,t,l)},\end{aligned}$$

which implies $\psi(\pi,t,k)\leq\psi(\pi,t,l)$.

(ii) Suppose $p_{22}\geq p_{12}$. Clearly, if $\psi(\pi,t,k)=0$, then $\psi(\pi,t,k)\leq\psi(\hat{\pi},t,k)$. Else if $\psi(\pi,t,k)>0$, because $\pi(p_{21}-p_{11})\geq\hat{\pi}(p_{21}-p_{11})$ and $\pi(p_{22}-p_{12})\leq\hat{\pi}(p_{22}-p_{12})$,

$$\begin{aligned}\frac{1-\psi(\pi,t,k)}{\psi(\pi,t,k)}&=\left(\frac{(1-\pi)p_{11}+\pi p_{21}}{(1-\pi)p_{12}+\pi p_{22}}\right)\left(\frac{q_{1k}(t+1)}{q_{2k}(t+1)}\right)\\&\geq\left(\frac{(1-\hat{\pi})p_{11}+\hat{\pi}p_{21}}{(1-\hat{\pi})p_{12}+\hat{\pi}p_{22}}\right)\left(\frac{q_{1k}(t+1)}{q_{2k}(t+1)}\right)=\frac{1-\psi(\hat{\pi},t,k)}{\psi(\hat{\pi},t,k)},\end{aligned}$$

which implies $\psi(\pi,t,k)\leq\psi(\hat{\pi},t,k)$. The proof for the case $p_{22}\leq p_{12}$ is analogous.

(iii) By the TP_2 property of $Q(t+1)$, $\sum_{k\geq l}q_{1k}(t+1)\leq\sum_{k\geq l}q_{2k}(t+1)$ (see, e.g., Proposition 1 in Albright [1]). Therefore, if $p_{22}\geq(\leq)\,p_{12}$,

$$\begin{aligned}\sum_{k\geq l}(\sigma(k;\pi,t)-\sigma(k;\hat{\pi},t))&=(\pi-\hat{\pi})(p_{21}-p_{11})\sum_{k\geq l}q_{1k}(t+1)\\&\quad+(\pi-\hat{\pi})(p_{22}-p_{12})\sum_{k\geq l}q_{2k}(t+1)\\&=(\pi-\hat{\pi})(p_{22}-p_{12})\sum_{k\geq l}(q_{2k}(t+1)-q_{1k}(t+1))\leq(\geq)\,0.\end{aligned}$$

□

In the proof of Lemma 1, we will also make use of the following technical result (cf. Puterman [32], Lemma 4.7.2).

Proposition 2. *Let f and g be probability mass functions on $\mathcal{O}$ such that $\sum_{k\geq l}f(k)\leq\sum_{k\geq l}g(k)$ for all $l\in\mathcal{O}$. Then $\sum_{k\in\mathcal{O}}f(k)h(k)\leq\sum_{k\in\mathcal{O}}g(k)h(k)$ for all nondecreasing functions $h\colon\mathcal{O}\to\mathbb{R}$.*

Proof of Lemma 1. We assume without loss of generality that $Q(t)$ satisfies the condition of Proposition 1 for all $t\in\mathbb{N}_0$ (see Remark 3). The proof proceeds by induction on the iterates of the value iteration algorithm (see Puterman [32], Section 6.3). We denote by V_n the value function at the nth iteration, for all $n\in\mathbb{N}_0$. We let V_0 be defined as in (3). The successive value functions are obtained through the dynamic programming recursion $V_{n+1}(\pi,t)=\min_{a\in\mathcal{A}}H^a(\pi,t,V_n)$ for all $(\pi,t)\in\Omega$, for all $n\in\mathbb{N}_0$. Two cases must be distinguished: (a) $p_{22}\geq p_{12}$ and (b) $p_{22}<p_{12}$.

Case (a) $p_{22}\geq p_{12}$. We will prove that, in all iterations $n\in\mathbb{N}_0$,

$$V_n(\hat{\pi},t)-V_n(\pi,t)\geq(\hat{\pi}-\pi)c_r \tag{7}$$

for all $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$. In iteration 0, inequality (7) is valid because $(1-\beta(p_{22}-p_{12}))^{-1}c_d > c_r$ by condition (4). We make the induction hypothesis that (7) holds in iteration m. In iteration $m+1$, for all $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$,

$$\begin{aligned} H^C(\hat{\pi}, t, V_m) - H^C(\pi, t, V_m) = (\hat{\pi} - \pi)c_d &+ \beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) V_m(\psi(\hat{\pi}, t, k), t+1) \\ &- \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) V_m(\psi(\pi, t, k), t+1). \end{aligned} \tag{8}$$

The key to the induction step is to determine a lower bound on (8). For that, we bound the second term on the right-hand side:

$$\begin{aligned} &\beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) V_m(\psi(\hat{\pi}, t, k), t+1) \\ &\geq \beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) \big[V_m(\psi(\pi, t, k), t+1) + (\psi(\hat{\pi}, t, k) - \psi(\pi, t, k))c_r \big] \qquad (9) \\ &= \beta((1-\hat{\pi})p_{12} + \hat{\pi}p_{22})c_r + \beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) \big[V_m(\psi(\pi, t, k), t+1) - \psi(\pi, t, k)c_r \big] \\ &\geq \beta((1-\hat{\pi})p_{12} + \hat{\pi}p_{22})c_r + \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) \big[V_m(\psi(\pi, t, k), t+1) - \psi(\pi, t, k)c_r \big] \qquad (10) \\ &= (\hat{\pi} - \pi)\beta(p_{22} - p_{12})c_r + \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) V_m(\psi(\pi, t, k), t+1). \end{aligned}$$

Inequality (9) follows from Proposition 1(ii) and the induction hypothesis. Proposition 2 can be applied to obtain inequality (10), using Proposition 1(iii) and the fact that, by Proposition 1(i) and the induction hypothesis, $V_m(\psi(\pi, t, k), t+1) - \psi(\pi, t, k)c_r$ is nondecreasing in k. Thus, we have

$$H^C(\hat{\pi}, t, V_m) - H^C(\pi, t, V_m) \geq (\hat{\pi} - \pi)\left(c_d + \beta(p_{22} - p_{12})c_r\right). \tag{11}$$

Using that $c_d + \beta(p_{22} - p_{12})c_r > c_r$ by condition (4), we can complete the induction step. For all $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$,

$$\begin{aligned} V_{m+1}(\hat{\pi}, t) - V_{m+1}(\pi, t) &= \min_{a \in \mathcal{A}} H^a(\hat{\pi}, t, V_m) - \min_{a \in \mathcal{A}} H^a(\pi, t, V_m) \\ &\geq \min_{a \in \mathcal{A}} \left(H^a(\hat{\pi}, t, V_m) - H^a(\pi, t, V_m) \right) \\ &\geq \min\{(\hat{\pi} - \pi)\left(c_d + \beta(p_{22} - p_{12})c_r\right), (\hat{\pi} - \pi)c_r, (\hat{\pi} - \pi)c_r\} \\ &= (\hat{\pi} - \pi)c_r, \end{aligned}$$

showing that (7) holds in iteration $m+1$. By induction, we conclude that (7) holds in all iterations $n \in \mathbb{N}_0$, and the result follows by noting that $V^* = \lim_{n \to \infty} V_n$.

Case (b) $p_{22} < p_{12}$. We repeat the steps taken in case (a), but for the induction to go through, we now need to strengthen the induction hypothesis by additionally propagating an upper bound on $V_n(\hat{\pi}, t) - V_n(\pi, t)$. We will prove that, in all iterations $n \in \mathbb{N}_0$,

$$(\hat{\pi} - \pi)c_r \leq V_n(\hat{\pi}, t) - V_n(\pi, t) \leq (\hat{\pi} - \pi)(c_d + \beta(p_{22} - p_{12})c_r) \tag{12}$$

for all $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$. Note that the base case (i.e., (12) holds in iteration 0) is still obtained directly from condition (4): if $p_{22} < p_{12}$, then by applying this condition twice, we obtain

$$c_r < (1-\beta(p_{22}-p_{12}))^{-1} c_d = c_d + \beta(p_{22}-p_{12})(1-\beta(p_{22}-p_{12}))^{-1} c_d \leq c_d + \beta(p_{22}-p_{12}) c_r.$$

We make the induction hypothesis that (12) holds in iteration m. In iteration $m+1$, we wish to bound $H^C(\hat{\pi}, t, V_m) - H^C(\pi, t, V_m)$ for all $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$. When we apply the same method as in case (a), the inequalities (9) and (10) reverse in accordance with the sign changes in parts (ii) and (iii) of Proposition 1. This yields an upper bound

$$H^C(\hat{\pi}, t, V_m) - H^C(\pi, t, V_m) \leq (\hat{\pi} - \pi)\left(c_d + \beta(p_{22}-p_{12}) c_r\right).$$

A lower bound can be obtained by a similar derivation, with the difference that (again, as a consequence of the sign changes in Proposition 1) the upper bound of the induction hypothesis must be used. We get

$$\begin{aligned}
&\beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) V_m(\psi(\hat{\pi}, t, k), t+1) \\
&\geq \beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) \big[V_m(\psi(\pi, t, k), t+1) + (\psi(\hat{\pi}, t, k) - \psi(\pi, t, k))(c_d + \beta(p_{22}-p_{12}) c_r) \big] \\
&= \beta((1-\hat{\pi})p_{12} + \hat{\pi} p_{22})(c_d + \beta(p_{22}-p_{12}) c_r) \\
&\qquad + \beta \sum_{k \in \mathcal{O}} \sigma(k; \hat{\pi}, t) \big[V_m(\psi(\pi, t, k), t+1) - \psi(\pi, t, k)(c_d + \beta(p_{22}-p_{12}) c_r) \big] \\
&\geq \beta((1-\hat{\pi})p_{12} + \hat{\pi} p_{22})(c_d + \beta(p_{22}-p_{12}) c_r) \\
&\qquad + \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) \big[V_m(\psi(\pi, t, k), t+1) - \psi(\pi, t, k)(c_d + \beta(p_{22}-p_{12}) c_r) \big] \\
&= (\hat{\pi} - \pi)\beta(p_{22}-p_{12})(c_d + \beta(p_{22}-p_{12}) c_r) + \beta \sum_{k \in \mathcal{O}} \sigma(k; \pi, t) V_m(\psi(\pi, t, k), t+1);
\end{aligned}$$

therefore,

$$H^C(\hat{\pi}, t, V_m) - H^C(\pi, t, V_m) \geq (\hat{\pi} - \pi)\left[c_d + \beta(p_{22}-p_{12})(c_d + \beta(p_{22}-p_{12}) c_r)\right].$$

To complete the induction step, we notice that by condition (4),

$$\begin{aligned}
c_d + \beta(p_{22}-p_{12})(c_d + \beta(p_{22}-p_{12}) c_r) &= (1 + \beta(p_{22}-p_{12})) c_d + (\beta(p_{22}-p_{12}))^2 c_r \\
&> (1 + \beta(p_{22}-p_{12}))(1 - \beta(p_{22}-p_{12})) c_r + (\beta(p_{22}-p_{12}))^2 c_r \\
&= c_r,
\end{aligned}$$

so $V_{m+1}(\hat{\pi}, t) - V_{m+1}(\pi, t) \geq (\hat{\pi} - \pi) c_r$ follows in the same way as in part (a). Furthermore, using

that $c_r < c_d + \beta(p_{22} - p_{12})c_r$ by condition (4), we have

$$\begin{aligned} V_{m+1}(\hat{\pi}, t) - V_{m+1}(\pi, t) &= \min_{a \in \mathcal{A}} H^a(\hat{\pi}, t, V_m) - \min_{a \in \mathcal{A}} H^a(\pi, t, V_m) \\ &\leq \max_{a \in \mathcal{A}} \left(H^a(\hat{\pi}, t, V_m) - H^a(\pi, t, V_m) \right) \\ &\leq \max\{(\hat{\pi} - \pi)\left(c_d + \beta(p_{22} - p_{12})c_r\right), (\hat{\pi} - \pi)c_r, (\hat{\pi} - \pi)c_r\} \\ &= (\hat{\pi} - \pi)\left(c_d + \beta(p_{22} - p_{12})c_r\right). \end{aligned}$$

This shows that (12) holds in iteration $m + 1$. By induction, we conclude that (12) holds in all iterations $n \in \mathbb{N}_0$, and the result follows by noting that $V^* = \lim_{n \to \infty} V_n$. □

For the proof of Lemma 2, it is useful to recall a central result in the theory of POMDPs by Smallwood and Sondik [37]. They consider a general POMDP formulation where rewards are maximized and show that the total expected discounted reward over a finite horizon is piecewise linear and convex in the information state. For our model, where the objective is to minimize the total expected discounted cost and the information state contains a completely observable variable, their result implies the following property of the value iteration algorithm we have used in the proof of Lemma 1.

Proposition 3. *If $V_0(\pi, t)$ is piecewise linear and concave in $\pi \in [0, 1]$ for all $t \in \mathbb{N}_0$, then in all iterations $n \in \mathbb{N}_0$, $V_n(\pi, t)$ is piecewise linear and concave in $\pi \in [0, 1]$ for all $t \in \mathbb{N}_0$.*

Proof of Lemma 2. As in the proof of Lemma 1, we will be using induction on the iterations of the value iteration algorithm. Again, we let V_0 be defined as in (3). (Note that V_0 satisfies the condition of Proposition 3.) We will prove that, in all iterations $n \in \mathbb{N}_0$,

$$V_n(\pi, t) \leq V_n(\pi, t + 1) \tag{13}$$

for all $(\pi, t) \in \Omega$. Observe that in iteration 0, (13) holds with equality. We make the induction hypothesis that (13) holds in iteration m. In the induction step, we start by showing that $H^C(\pi, t, V_m) \leq H^C(\pi, t + 1, V_m)$ for all $(\pi, t) \in \Omega$. Consider a fixed $(\pi, t) \in \Omega$ and let $l \in \mathcal{O}$ such that $\sigma(l; \pi, t + 1) > 0$. The relation $Q(t + 1) \succeq_B Q(t + 2)$ allows writing $\psi(\pi, t + 1, l)$ as a convex combination of $\psi(\pi, t, k)$, $k \in \mathcal{O}$. Specifically,

$$\psi(\pi, t + 1, l) = \sum_{k \in \mathcal{O}} \left(\frac{\sigma(k; \pi, t) x_{kl}(t + 1)}{\sigma(l; \pi, t + 1)} \right) \psi(\pi, t, k),$$

where

$$\sum_{k \in \mathcal{O}} \left(\frac{\sigma(k; \pi, t) x_{kl}(t + 1)}{\sigma(l; \pi, t + 1)} \right) = 1.$$

With the concavity of V_m (see Proposition 3), by Jensen's inequality this implies

$$\sum_{k \in \mathcal{O}} \left(\frac{\sigma(k; \pi, t) x_{kl}(t + 1)}{\sigma(l; \pi, t + 1)} \right) V_m(\psi(\pi, t, k), t + 1) \leq V_m(\psi(\pi, t + 1, l), t + 1).$$

Upon rewriting and summing over all $l \in \mathcal{O}$, including values of l such that $\sigma(l;\pi,t+1) = \sum_{k\in\mathcal{O}} \sigma(k;\pi,t)x_{kl}(t+1) = 0$, we get

$$\sum_{k,l\in\mathcal{O}} \sigma(k;\pi,t)x_{kl}(t+1)V_m(\psi(\pi,t,k),t+1) \leq \sum_{l\in\mathcal{O}} \sigma(l;\pi,t+1)V_m(\psi(\pi,t+1,l),t+1). \tag{14}$$

The immediate cost does not depend on t; therefore, from (14),

$$\begin{aligned} H^C(\pi,t,V_m) &= \pi c_d + \beta \sum_{k\in\mathcal{O}} \sigma(k;\pi,t)V_m(\psi(\pi,t,k),t+1) \\ &= \pi c_d + \beta \sum_{k,l\in\mathcal{O}} \sigma(k;\pi,t)x_{kl}(t+1)V_m(\psi(\pi,t,k),t+1) \\ &\leq \pi c_d + \beta \sum_{l\in\mathcal{O}} \sigma(l;\pi,t+1)V_m(\psi(\pi,t+1,l),t+1) \\ &\leq \pi c_d + \beta \sum_{l\in\mathcal{O}} \sigma(l;\pi,t+1)V_m(\psi(\pi,t+1,l),t+2) \qquad (15) \\ &= H^C(\pi,t+1,V_m), \end{aligned}$$

where inequality (15) holds by the induction hypothesis. Because, in particular, the above shows that $H^C(0,t,V_m) \leq H^C(0,t+1,V_m)$ for all $t \in \mathbb{N}_0$, it follows that $H^S(\pi,t,V_m) \leq H^S(\pi,t+1,V_m)$ for all $(\pi,t) \in \Omega$. Further, clearly, $H^R(\pi,t,V_m) \leq H^R(\pi,t+1,V_m)$ holds with equality for all $(\pi,t) \in \Omega$. To complete the induction step, we combine the inequalities for each action into

$$V_{m+1}(\pi,t) = \min_{a\in\mathcal{A}} H^a(\pi,t,V_m) \leq \min_{a\in\mathcal{A}} H^a(\pi,t+1,V_m) = V_{m+1}(\pi,t+1)$$

for all $(\pi,t) \in \Omega$. This establishes that (13) holds in iteration $m+1$. By induction, we conclude that (13) holds in all iterations $n \in \mathbb{N}_0$, and the result follows by noting that $V^* = \lim_{n\to\infty} V_n$. □

We will give two lemmas with results on the optimal policy under condition (4), which together lead to the conclusion of Theorem 2. The first lemma provides results to establish, for all $t \in \mathbb{N}_0$, the existence of a threshold $\pi^*(t) \in [0,1)$ such that $\delta^*(\pi,t) = C$ if and only if $\pi \leq \pi^*(t)$.

Lemma 3. *If condition (4) holds, then*

(i) if $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$, then $\delta^*(\hat{\pi},t) = C$ implies $\delta^*(\pi,t) = C$, for all $t \in \mathbb{N}_0$;

(ii) $\delta^*(0,t) = C$, for all $t \in \mathbb{N}_0$;

(iii) $\delta^*(1,t) \in \{S,R\}$, for all $t \in \mathbb{N}_0$.

Proof.

(i) Let $\pi, \hat{\pi} \in [0,1]$ such that $\pi \leq \hat{\pi}$ and $t \in \mathbb{N}_0$. From Lemma 1, we can derive that

$$\begin{aligned} \min_{a\in\{S,R\}} H^a(\pi,t,V^*) - H^C(\pi,t,V^*) &= \min_{a\in\{S,R\}} H^a(\hat{\pi},t,V^*) - (\hat{\pi}-\pi)c_r - H^C(\pi,t,V^*) \\ &\geq \min_{a\in\{S,R\}} H^a(\hat{\pi},t,V^*) - H^C(\hat{\pi},t,V^*). \end{aligned}$$

It follows that if $\delta^*(\hat{\pi},t) = C$, then also $\delta^*(\pi,t) = C$.

(ii) For $t = 0$, because it is clear that $H^C(0,0,V^*) < H^S(0,0,V^*) < H^R(0,0,V^*)$, the result $\delta^*(0,t) = C$ is immediate. So let us consider $t > 0$. Again, it is clear that $H^C(0,t,V^*) < H^S(0,t,V^*)$ and, therefore, $\delta^*(0,t) \neq S$. To also prove that $\delta^*(0,t) \neq R$, we proceed by establishing inequalities with respect to $H^C(0,t,V^*)$ and $H^R(0,t,V^*)$. We have

$$\begin{aligned}
H^C(0,t,V^*) &= \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)V^*(\psi(0,t,k),t+1) \\
&\leq \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)H^R(\psi(0,t,k),t+1,V^*) \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)(c_s + c_i + \psi(0,t,k)c_r + H^C(0,0,V^*)) \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)(c_s + c_i + \psi(0,t,k)c_r + V^*(0,0)) \qquad (16) \\
&= \beta(c_s + c_i + p_{12}c_r + V^*(0,0)),
\end{aligned}$$

where Equation (16) is valid because $\delta^*(0,0) = C$, and on the other hand,

$$\begin{aligned}
H^R(0,t,V^*) &= c_s + c_i + \beta \sum_{k\in\mathcal{O}} \sigma(k;0,0)V^*(\psi(0,0,k),1) \\
&\geq c_s + c_i + \beta \sum_{k\in\mathcal{O}} \sigma(k;0,0)V^*(\psi(0,0,k),0) \qquad (17) \\
&\geq c_s + c_i + \beta \sum_{k\in\mathcal{O}} \sigma(k;0,0)(\psi(0,0,k)c_r + V^*(0,0)) \qquad (18) \\
&= c_s + c_i + \beta(p_{12}c_r + V^*(0,0)),
\end{aligned}$$

where Equations (17) and (18) are obtained using Lemmas 2 and 1, respectively. These inequalities imply

$$H^C(0,t,V^*) \leq \beta(c_s + c_i + p_{12}c_r + V^*(0,0)) < c_s + c_i + \beta(p_{12}c_r + V^*(0,0)) \leq H^R(0,t,V^*).$$

We conclude that $\delta^*(0,t) = C$.

(iii) Let $t \in \mathbb{N}_0$, and let the value function V be defined as in (3). We first show that $\max_{\pi\in[0,1]}(V(\pi,t) - V^*(\pi,t)) = V(1,t) - V^*(1,t)$. The maximum exists and is attained at either $\pi = 0$ or $\pi = 1$ because $V(\pi,t)$ is linear in π and, by Proposition 3 and convergence of the value iteration algorithm, $V^*(\pi,t)$ is concave in π. To prove that the maximum is attained at $\pi = 1$, we argue by contradiction. Suppose that $\max_{\pi\in[0,1]}(V(\pi,t) - V^*(\pi,t)) = V(0,t) - V^*(0,t)$. Given that we have established in the proof of Theorem 1 that $V(0,t) = H^C(0,t,V)$ and, by part (ii) of

this lemma, $V^*(0,t) = H^C(0,t,V^*)$, we have

$$\begin{aligned}
V(0,t) - V^*(0,t) &= H^C(0,t,V) - H^C(0,t,V^*) \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)\left[V(\psi(0,t,k),t+1) - V^*(\psi(0,t,k),t+1)\right] \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)\left[V(\psi(0,t,k),t) - V^*(\psi(0,t,k),t+1)\right] \\
&\leq \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)\left[V(\psi(0,t,k),t) - V^*(\psi(0,t,k),t)\right] \qquad (19) \\
&\leq \beta \sum_{k\in\mathcal{O}} \sigma(k;0,t)\left[V(0,t) - V^*(0,t)\right] \qquad (20) \\
&= \beta(V(0,t) - V^*(0,t)),
\end{aligned}$$

where inequality (19) is obtained by Lemma 2 and inequality (20) is a consequence of the assumption that $\max_{\pi\in[0,1]}(V(\pi,t) - V^*(\pi,t)) = V(0,t) - V^*(0,t)$. Because $\beta < 1$, from the above we get $V(0,t) - V^*(0,t) \leq 0$, which would mean $\max_{\pi\in[0,1]}(V(\pi,t) - V^*(\pi,t)) \leq 0$. However, this gives a contradiction with $V(1,t) - V^*(1,t) > 0$ as has been established in the proof of Theorem 1. Hence, it must hold that $\max_{\pi\in[0,1]}(V(\pi,t) - V^*(\pi,t)) = V(1,t) - V^*(1,t)$. We are now ready to prove that $\delta^*(1,t) \in \{S,R\}$. Again, we argue by contradiction. Suppose that $\delta^*(1,t) = C$. Following the same steps as above, we get

$$\begin{aligned}
V(1,t) - V^*(1,t) &= H^C(1,t,V) - H^C(1,t,V^*) \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;1,t)\left[V(\psi(1,t,k),t+1) - V^*(\psi(1,t,k),t+1)\right] \\
&= \beta \sum_{k\in\mathcal{O}} \sigma(k;1,t)\left[V(\psi(1,t,k),t) - V^*(\psi(1,t,k),t+1)\right] \\
&\leq \beta \sum_{k\in\mathcal{O}} \sigma(k;1,t)\left[V(\psi(1,t,k),t) - V^*(\psi(1,t,k),t)\right] \\
&\leq \beta \sum_{k\in\mathcal{O}} \sigma(k;1,t)\left[V(1,t) - V^*(1,t)\right] \\
&= \beta(V(1,t) - V^*(1,t)).
\end{aligned}$$

As before, because this would mean that $V(1,t) - V^*(1,t) \leq 0$, we arrive at a contradiction with $V(1,t) - V^*(1,t) > 0$ as has been established in the proof of Theorem 1. We conclude that $\delta^*(1,t) \in \{S,R\}$. □

The results in the second lemma are to show that if there is an information state in which full inspection with sensor replacement is optimal, then there exists a threshold $t^* \in \mathbb{N}_0$ such that for all $\pi \in [0,1]$, $\delta^*(\pi,t) \in \{C,S\}$ if $t \leq t^*$ and $\delta^*(\pi,t) \in \{C,R\}$ if $t > t^*$, and $\pi^*(t)$ is nonincreasing in t for $t > t^*$.

Lemma 4. *If condition (4) holds, then*

(i) $\delta^*(1,t) = R$ *implies* $\delta^*(1,t+1) = R$, *for all* $t \in \mathbb{N}_0$;

(ii) *for all* $t \in \mathbb{N}_0$, *if* $\delta^*(1,t) = S$, *then* $\delta^*(\pi,t) \in \{C,S\}$ *for all* $\pi \in [0,1]$, *and conversely, if* $\delta^*(1,t) = R$, *then* $\delta^*(\pi,t) \in \{C,R\}$ *for all* $\pi \in [0,1]$;

(iii) $\delta^*(1,0) = S$;

(iv) $\delta^*(\pi,t) = R$ *implies* $\delta^*(\pi,t+1) = R$, *for all* $(\pi,t) \in \Omega$.

Proof.

(i) Let $t \in \mathbb{N}_0$. From Lemma 2, we can derive that

$$\begin{aligned} H^R(1,t,V^*) - H^S(1,t,V^*) &\geq H^R(1,t,V^*) - H^S(1,t+1,V^*) \\ &= H^R(1,t+1,V^*) - H^S(1,t+1,V^*). \end{aligned}$$

Because Lemma 3(iii) indicates that $\delta^*(1,t) \in \{S,R\}$ and $\delta^*(1,t+1) \in \{S,R\}$, it follows that if $\delta^*(1,t) = R$, then also $\delta^*(1,t+1) = R$.

(ii) The result holds because $H^R(\pi,t,V^*) - H^S(\pi,t,V^*)$ is constant in π, for all $t \in \mathbb{N}_0$.

(iii) Lemma 3(iii) gives $\delta^*(1,0) \in \{S,R\}$, and it is obvious that $H^S(1,0,V^*) < H^R(1,0,V^*)$. Hence, $\delta^*(1,0) = S$.

(iv) Let $(\pi,t) \in \Omega$. If $\delta^*(\pi,t) = R$, then parts (i) and (ii) of this lemma guarantee that $\delta^*(\pi,t+1) \in \{C,R\}$. In addition, we can derive from Lemma 2 that

$$\begin{aligned} H^R(\pi,t,V^*) - H^C(\pi,t,V^*) &\geq H^R(\pi,t,V^*) - H^C(\pi,t+1,V^*) \\ &= H^R(\pi,t+1,V^*) - H^C(\pi,t+1,V^*). \end{aligned}$$

It follows that if $\delta^*(\pi,t) = R$, then also $\delta^*(\pi,t+1) = R$. □

Proof of Theorem 2. The result follows directly from Lemmas 3 and 4. □

Proof of Theorem 3. For all $(\pi,t) \in \widetilde{\Omega} \setminus \bar{\Omega}$, it holds that

$$\begin{aligned} H_L(\pi,t,V^*) &= (\lceil \pi z \rceil - \pi z) H(\lfloor \pi z \rfloor / z, t, V^*) + (\pi z - \lfloor \pi z \rfloor) H(\lceil \pi z \rceil / z, t, V^*) \\ &= (\lceil \pi z \rceil - \pi z) V^*(\lfloor \pi z \rfloor / z, t) + (\pi z - \lfloor \pi z \rfloor) V^*(\lceil \pi z \rceil / z, t) \\ &\leq V^*(\pi,t) \qquad (21) \\ &\leq V^*(\lceil \pi z \rceil / z, t) - (\lceil \pi z \rceil / z - \pi) c_r \qquad (22) \\ &= H(\lceil \pi z \rceil / z, t, V^*) - (\lceil \pi z \rceil / z - \pi) c_r \\ &= H_U(\pi,t,V^*). \end{aligned}$$

The concavity of $V^*(\pi,t)$ in π (which follows from Proposition 3 and convergence of the value iteration algorithm) implies inequality (21), while inequality (22) is valid because of Lemma 1. Since for $(\pi,t) \in \bar{\Omega}$,

$$H_L(\pi,t,V^*) = H(\pi,t,V^*) = V^*(\pi,t) = H(\pi,t,V^*) = H_U(\pi,t,V^*),$$

this means that we have $V^*(\pi,t) \geq H_L(\pi,t,V^*)$ and $V^*(\pi,t) \leq H_U(\pi,t,V^*)$ for all $(\pi,t) \in \widetilde{\Omega}$. Hence, using parts (a) and (b) of Theorem 6.2.2 in Puterman [32], we conclude that $V^*(\pi,t) \geq V_L(\pi,t)$ and $V^*(\pi,t) \leq V_U(\pi,t)$ for all $(\pi,t) \in \widetilde{\Omega}$. □

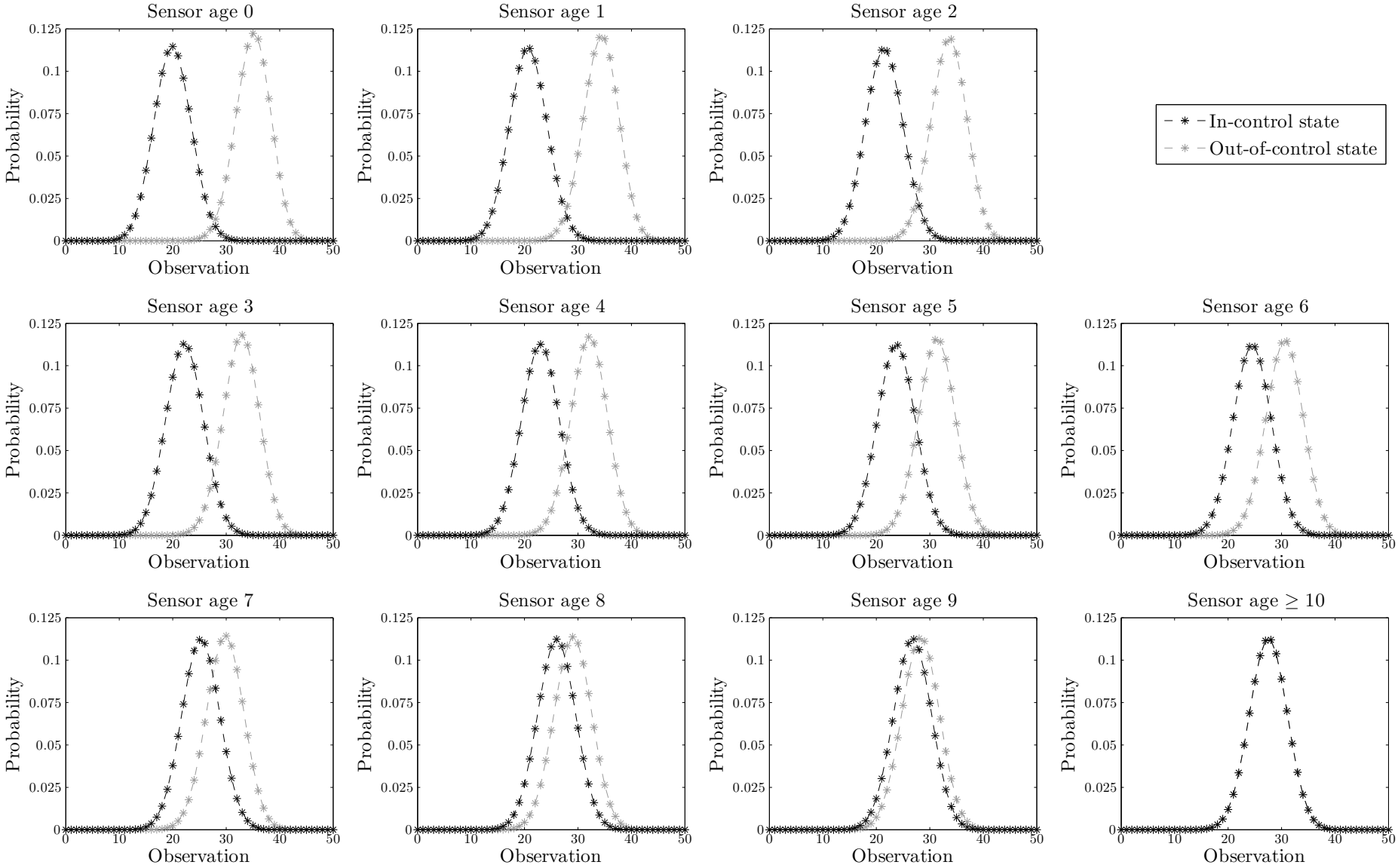

Figure 1: Probability mass function of the sensor observation as a function of the sensor age in Example 1.

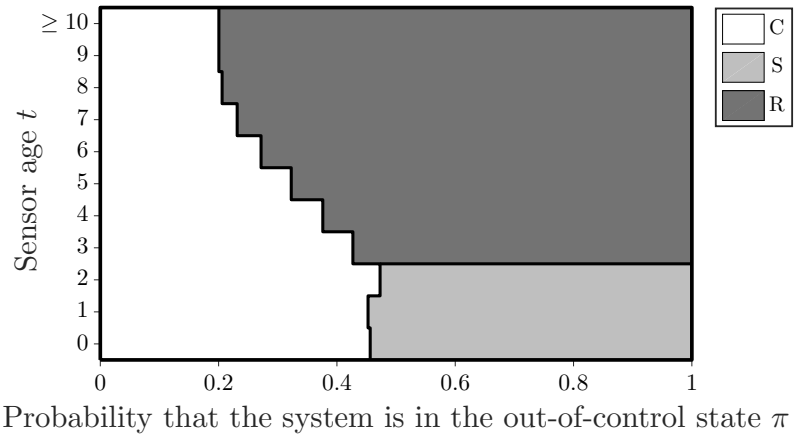

Figure 2: Optimal policy for Example 1. The threshold values are $\bar{\pi}^*(0) = 0.456$, $\bar{\pi}^*(1) = 0.453$, $\bar{\pi}^*(2) = 0.473$, $\bar{\pi}^*(3) = 0.427$, $\bar{\pi}^*(4) = 0.376$, $\bar{\pi}^*(5) = 0.323$, $\bar{\pi}^*(6) = 0.272$, $\bar{\pi}^*(7) = 0.231$, $\bar{\pi}^*(8) = 0.206$, $\bar{\pi}^*(9) = 0.200$, $\bar{\pi}^*(10) = 0.200$, and $\bar{t}^* = 2$.

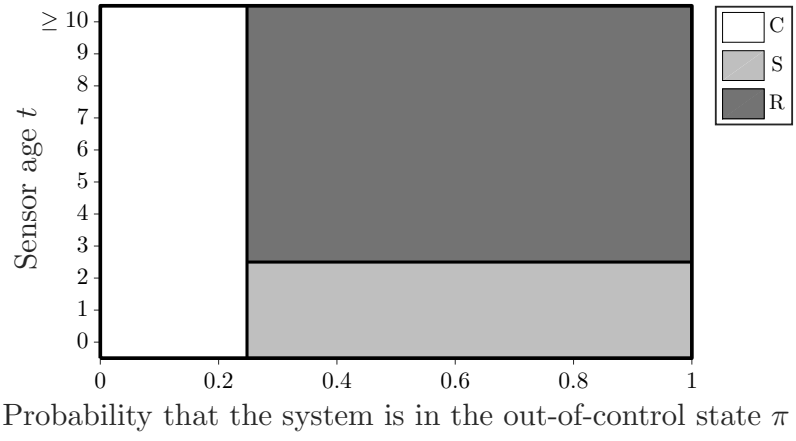

Figure 3: Heuristic policy H_1 for Example 1. The threshold values are $\bar{\pi}^{H_1} = 0.248$ and $\bar{t}^{H_1} = 2$.

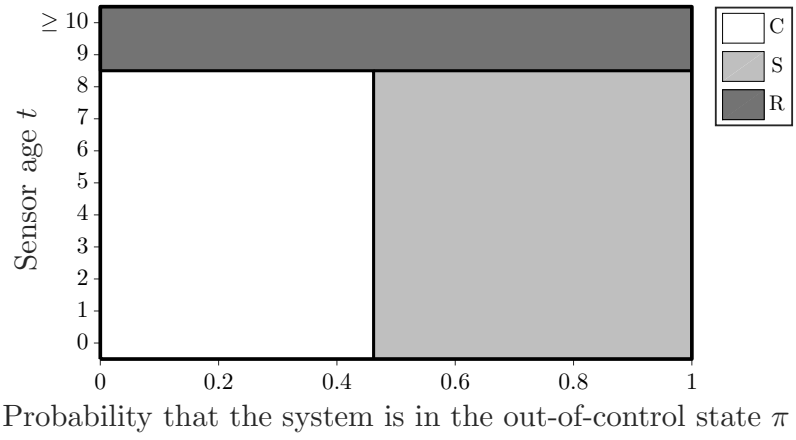

Figure 4: Heuristic policy H_2 for Example 1. The threshold values are $\bar{\pi}^{H_2} = 0.462$ and $\bar{t}^{H_2} = 8$.

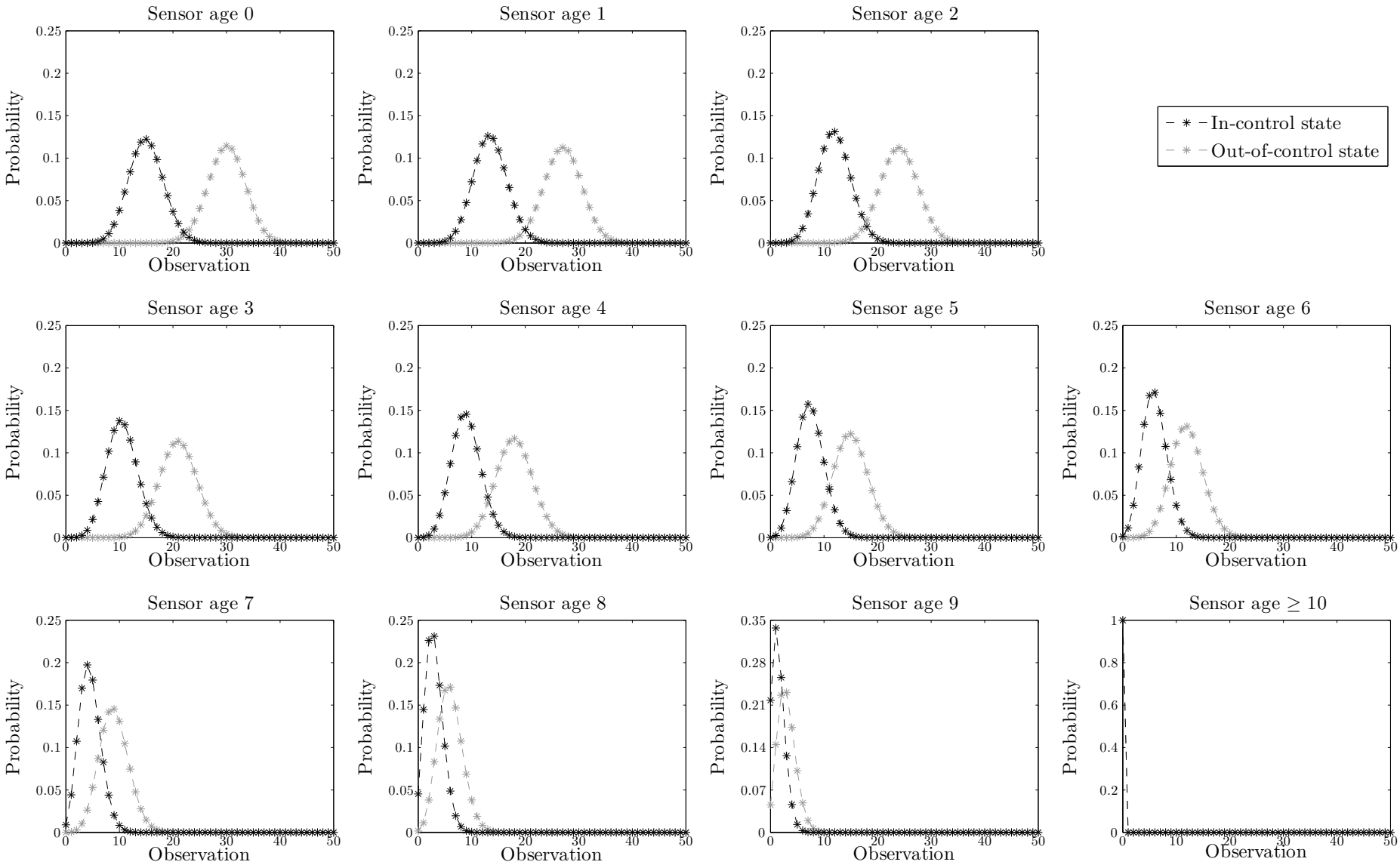

Figure 5: Probability mass function of the sensor observation as a function of the sensor age in Example 2.

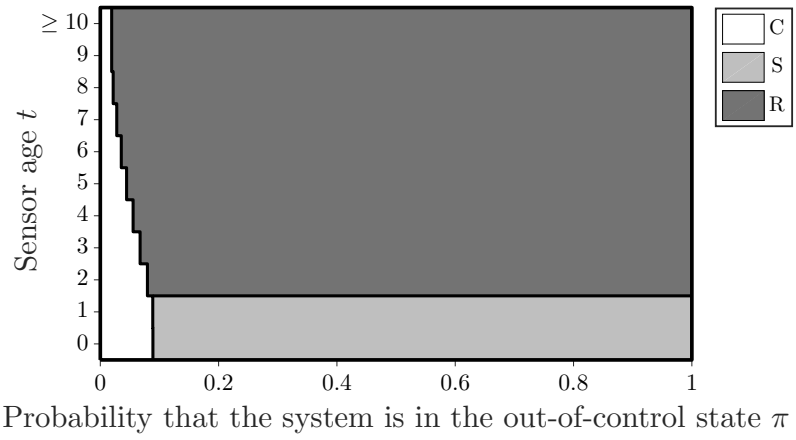

Figure 6: Optimal policy for Example 2. The threshold values are $\bar{\pi}^*(0) = 0.089$, $\bar{\pi}^*(1) = 0.089$, $\bar{\pi}^*(2) = 0.080$, $\bar{\pi}^*(3) = 0.067$, $\bar{\pi}^*(4) = 0.056$, $\bar{\pi}^*(5) = 0.044$, $\bar{\pi}^*(6) = 0.036$, $\bar{\pi}^*(7) = 0.028$, $\bar{\pi}^*(8) = 0.022$, $\bar{\pi}^*(9) = 0.019$, $\bar{\pi}^*(10) = 0.019$, and $\bar{t}^* = 1$.

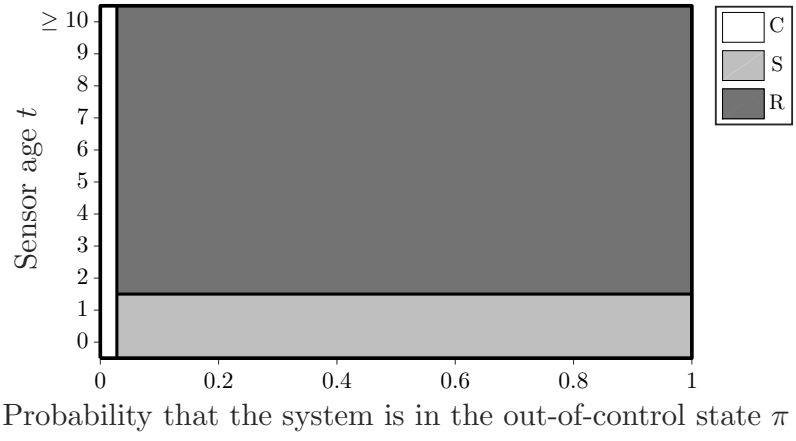

Figure 7: Heuristic policy H_1 for Example 2. The threshold values are $\bar{\pi}^{H_1} = 0.028$ and $\bar{t}^{H_1} = 1$.

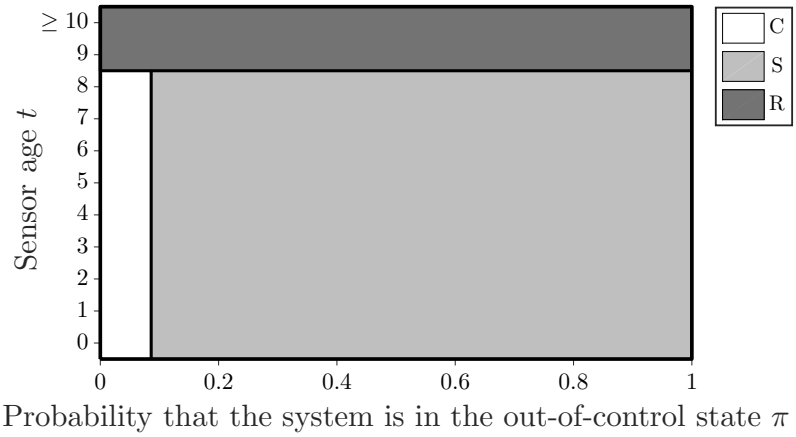

Figure 8: Heuristic policy H_2 for Example 2. The threshold values are $\bar{\pi}^{H_2} = 0.086$ and $\bar{t}^{H_2} = 8$.